



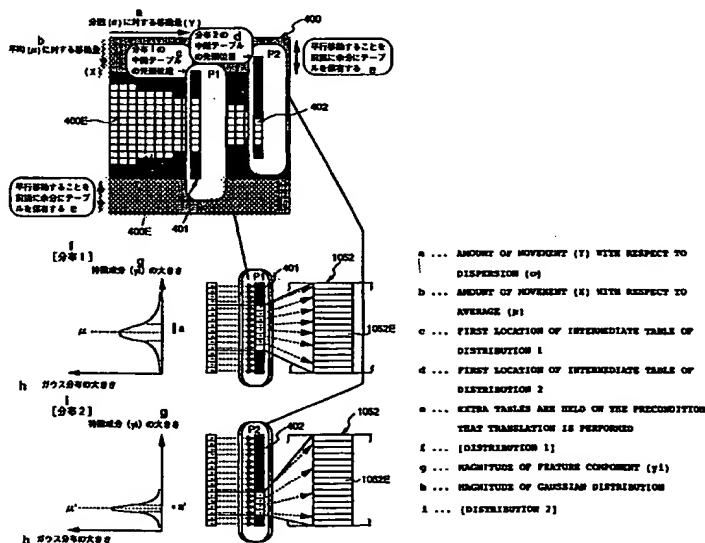
(51) 国際特許分類6 G10L 3/00	A1	(11) 国際公開番号 WO00/46788
		(43) 国際公開日 2000年8月10日(10.08.00)
(21) 国際出願番号 PCT/JP99/00493	(81) 指定国 AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, 欧州特許 (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG), ARIPO特許 (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), ユーラシア特許 (AM, AZ, BY, KC, KZ, MD, RU, TJ, TM)	
(22) 国際出願日 1999年2月5日(05.02.99)	添付公開書類 国際調査報告書	
(71) 出願人 (米国を除くすべての指定国について) 株式会社 日立製作所(HITACHI, LTD.)(JP/JP) 〒101-8010 東京都千代田区神田駿河台四丁目6番地 Tokyo, (JP)		
(72) 発明者 ; および		
(75) 発明者 / 出願人 (米国についてのみ) 中川智仁(NAKAGAWA, Motohito)(JP/JP) 前島英雄(MAEJIMA, Hideo)(JP/JP) 〒187-8588 東京都小平市上水本町五丁目20番1号 株式会社 日立製作所 半導体事業本部内 Tokyo, (JP)		
(74) 代理人 弁理士 玉村静世(TAMAMURA, Shizuyo) 〒271-0092 千葉県松戸市松戸1333番地1 コスモ松戸ステーションスクエア209号 Chiba, (JP)		

(54)Title: DATA PROCESSING SYSTEM

(54)発明の名称 データ処理システム

**(57) Abstract**

The output probability of a mixture Gaussian HMM is given by a function, such as a mixture multi-dimensional Gaussian distribution. A mixture multi-dimensional Gaussian distribution is the sum of multi-dimensional Gaussian distributions, and a multi-dimensional Gaussian distribution is the product of one-dimensional Gaussian distributions of respective feature components. The values of distribution relating to the representative dispersions and averages of one-dimensional distributions are listed in a numeric value table (1052). For the feature component of a feature vector, which is an object to be recognized, linear scalar quantization is adopted, and by using the quantized value as index, information in intermediate tables (401, 402) is referred to. The intermediate tables are prepared by extraction from a global table (400). The global table contains X-direction arrays in the Y direction, each of which is an array of address information representing the locations of numeric values of each one-dimensional Gaussian distribution listed in the numeric value table. The value of the average ( $\mu$ ) is taken into consideration. Tables are extracted from a global table, and a numeric quantized value is used as index according to the first



## (57)要約

混合ガウスHMMの出力確率は、混合多次元ガウス分布のような関数で与えられる。混合多次元ガウス分布は、多次元ガウス分布の和になり、多次元ガウス分布は特徴成分毎の1次元ガウス分布の積になる。1次元ガウス分布の代表的な分散及び平均に係る分布の数値は数値テーブル(1052)が保有する。認識対象とされる特徴ベクトルの特徴成分に対して線形スカラ量子化を採用し、その量子化値をインデックスとして中間テーブル(401, 402)上の情報を参照する。中間テーブルは、グローバルテーブル(400)から抽出する。グローバルテーブルは、数値テーブル上の1次元ガウス分布の数値の所在を示すアドレス情報の配列になっているX方向の配列をY方向に複数組有する。X方向配列の選択は分散( $\sigma$ )の値を考慮し、配列のX方向先頭位置の選択には平均( $\mu$ )の値を考慮する。これによって、グローバルテーブルから中間テーブルを抽出し、その先頭位置から前記量子化値をインデックスとする位置のアドレス情報を用いて数値テーブルを参照する。

PCTに基づいて公開される国際出願のパンフレット第一頁に掲載されたPCT加盟国を同定するために使用されるコード(参考情報)

AE アラブ首長国連邦	DM ドミニカ	KZ カザフスタン	RU ロシア
AG アンティグア・バーブーダ	DZ アルジェリア	LC セントルシア	SD スーダン
AL アルバニア	EE エストニア	LI リヒテンシュタイン	SE スウェーデン
AM アルメニア	ES スペイン	LK スリ・ランカ	SG シンガポール
AT オーストリア	FI フィンランド	LR リベリア	SI スロヴェニア
AU オーストラリア	FR フランス	LS レソト	SK スロヴァキア
AZ アゼルバイジャン	GA ガボン	LT リトアニア	SL シェラ・レオネ
BA ボスニア・ヘルツェゴビナ	GB 英国	LU ルクセンブルグ	SN セネガル
BB バルバドス	GD グレナダ	LV ラトヴィア	SZ スワジランド
BE ベルギー	GE グルジア	MA モロッコ	TD チャード
BF ブルキナ・ファソ	GH ガーナ	MC モナコ	TG トーゴ
BG ブルガリア	GM ガンビア	MD モルドヴァ	TJ タジキスタン
BJ ベナン	GN ギニア	MG マダガスカル	TM トルクメニスタン
BR ブラジル	GR ギリシャ	MK マケドニア・旧ユーゴスラヴィア	TR トルコ
BY ベラルーシ	GW ギニア・ビサウ	共和国	TT トリニダード・トバゴ
CA カナダ	HR クロアチア	マリ	TZ タンザニア
CF 中央アフリカ	HU ハンガリー	MN モンゴル	UA ウクライナ
CG コンゴ	ID インドネシア	MR モーリタニア	UG ウガンダ
CH スイス	IE アイルランド	MW マラウイ	US 米国
CI コートジボアール	IL イスラエル	MX メキシコ	UZ ウズベキスタン
CM カメルーン	IN インド	MZ モザンビーク	VN ヲトナム
CN 中国	IS アイスランド	NE ニジェール	YU ユーゴスラヴィア
CR コスタ・リカ	IT イタリア	NL オランダ	ZA 南アフリカ共和国
CU キューバ	JP 日本	NO ノールウェー	ZW ジンバブエ
CY キプロス	KE ケニア	NZ ニュー・ジーランド	
CZ チェッコ	KG キルギスタン	PL ポーランド	
DE ドイツ	KP 北朝鮮	PT ポルトガル	
DK デンマーク	KR 韓国	RO ルーマニア	

## 明 細 書

## データ処理システム

## 5 技術分野

本発明は、混合ガウス分布を用いた連続分布型隠れマルコフモデル (Continuous Mixture Hidden Markov Models : HMM) を用いた音声認識技術、更にはそのための出力確率の演算技術に関し、例えば、音声認識のための演算処理を行うデータプロセッサを有し電池駆動される  
10 携帯情報端末装置に適用して有効な技術に関するものである。

## 背景技術

隠れマルコフモデルは、マルコフ過程 (時点  $t + 1$  の状態が時点  $n$  の状態によってのみ与えられる確率過程) で表現される状態遷移モデルである。この隠れマルコフモデルを音声認識技術に適用することができる。  
15 この音声認識技術の概要を解かりやすく説明する。認識対象音声を例えば  $10\text{ ms}$  のような部分区間 (フレーム) に分割し、フレーム毎に周波数スペクトルのような特徴ベクトルを抽出する。このとき、認識すべき音の連鎖をフレーム毎の状態の連鎖とみなす。個々の状態に対して特徴ベクトルに近似した音源を割当てられるように各状態を決定できれば、音声認識が実現されることになる。そのために、個々の状態が種々の音源に対して特徴ベクトルに匹敵する尤度 (もっともらしさ) を確率的に表す出力確率と、現在の状態が隣の状態に変化する確率を示す状態遷移確率とを採用し、前記状態毎の出力確率と状態遷移確率との積の各  
20 状態の総和が最も大きくなる状態の連鎖を以って、音声認識結果とすることができる。前記特徴ベクトル列から想定されるパターン毎に、フレ

一ム単位で状態遷移確率と出力確率とを積算するための演算量は膨大である。特に、出力確率は混合多次元ガウス分布によって与えられる。混合多次元ガウス分布は、例えば、「あ」という音素に対して年齢、性別などの個々の要素毎の確率的な分布を有することになり、各々の確率的な分布は特徴ベクトルの次数に応ずる多次元ガウス分布から成り、  
5 夫々の多次元ガウス分布は1次元のガウス分布を複合した確率分布とされる。したがって、混合多次元ガウス分布における混合数や次数が多いほど出力確率の演算に時間を要することになる。本発明者の試算に依れば、出力確率の計算負荷は音声認識処理全体の50～80%のように  
10 膨大になることがある。

この出力確率の演算を高速化するには、混合多次元ガウス分布に対して、計算する分布の範囲を削減することが有効である。例えば、特徴ベクトルを幾つかの標準的なパターンに対応させ（ベクトル量子化）、そのパターン毎に出力確率を定義する方法が可能である。この処理の場合、  
15 特徴空間を部分領域に分割して、その部分領域と計算する分布とを対応付けることになるが、このような特徴ベクトルと部分領域との対応には、ベクトル量子化を用いることができる。ベクトル量子化とは、特徴空間上の有限個の代表ベクトルを考え、特徴空間上の任意の点を、その点に最も近い代表ベクトルで近似表現する方法である。このようなベクトル  
20 量子化には、効率的な方法が幾つか提案されているが、基本的に、距離が最小となる代表ベクトルを選択するもので、混合分布を計算するのに比べて計算量は僅少ではあるが、それでも計算負荷が小さいとはいえない。

また、出力確率の演算の一部をテーブル化して演算速度を高速化することも可能である。この場合にも、そのテーブルをベクトル量子化によって構成することができる。しかし、ベクトル量子化して出力確率を対  
25

応させると、量子化誤差が大きくなり、認識性能が劣化する。

そこで、計算を各特徴次元における計算に分解して、各特徴次元を標準的なガウス分布のパターンに分割し、各々の計算結果をテーブル化することが考えられる。このような手法にスカラ量子化を採用する。スカ  
5 ラ量子化として、例えば、単一ガウス分布をテーブル化する手法がある。この場合、ベクトル量子化とは異なり、量子化誤差は僅少になる。

スカラ量子化として、非線形スカラ量子化を採用することができる。即ち、特徴ベクトルの特徴次数は数十次元におよび、混合多次元ガウス分布の各次元毎に、全ての単一ガウス分布をテーブル化することは効率的でないから、データテーブルの種類を少なくしようとするものである。  
10 混合ガウス分布のスカラ量子化において、各次元毎の関数は、単一の1次元正規分布（単一ガウス分布）であり、これを用いれば、出力確率の計算を簡略化することができる。特徴次数や混合毎に相異されるであろう1次元正規分布の相関は、夫々の分布の平均と分散が分かれば規定で  
15 きる。その相関を決めるために、特徴次数毎にパラメータを演算し、演算したパラメータと特徴ベクトルの特徴成分とを用いて、代表的に設けられている1次元正規分布の数値テーブルをアクセスする。このような非線形スカラ量子化を用いて数値テーブルをアクセスして、混合HMMの演算量を減らそうとする技術については、例えば、“ON THE USE OF  
20 SCALAR QUANTIZATION FOR FAST HMM COMPUTATION”, ICASSP 95, pp.213-216がある。

しかしながら、その手法は、テーブルアクセスのために必ず各特徴成分毎のパラメータ演算を行わなければならない、また、テーブルの参照においても、そのように演算されたパラメータを用いるアクセスは、  
25 テーブルに対して連続した配列のアクセスになるとは限らないので、テーブルを参照するためのアドレス演算も毎回乗算と加算が必要になる。

このような手間のかかるパラメータ演算を行わずに数値テーブルを参照できるようにするには、例えば、一般的な線形量子化を適用した線形スカラ量子化を行うようにすればよい。即ち、特徴を等間隔に量子化するようにする。例えば、量子化し易いように、単一ガウス分布のデータテーブルを2のN乗個に分割すれば、特徴成分の上位Nビットを抽出することによって、簡単に量子化することができる。線形スカラ量子化では、代表点は固定であるから、混合多次元ガウス分布に対して、線形スカラ量子化処理は、各フレーム毎に1回行えばよい。換言すれば、特徴次元毎に1回行えばよい。また、代表点はそのままインデックスに相当するから、数値テーブルにおける先頭アドレスと所望アドレスとの差（以下、オフセット）は、インデックス×データ長であり、これも全ての分布に共通であるから、そのような演算は1フレームにただ1回実行すれば良い。そして、必要な数値テーブルへのアクセスは、各数値テーブルのアドレスと、全ての特徴成分に共通したオフセットの和で計算できるから、結局、1回の加算と、2回のロード（先頭アドレスと数値データ）で実行される。

混合ガウス型HMMの出力確率の計算では、単一ガウス分布（含む対数系）に対する計算量の低減が重要になる。各特徴成分毎の単一ガウス分布の計算は、出力確率の計算において最も計算負荷の大きい部分であり、計算数は、全モデル数×混合数×特徴次元の数で表現されるため、些細な計算コストの増大が、全体の計算量の増大に直結する。この部分で、線形スカラ量子化は、テーブルのアクセス以外、全く計算が発生しないため、計算効率の観点からは非常に優れている。

しかしながら、線形スカラ量子化は、計算効率の観点から非常に高速であるが、固定された代表点に対して、各分布毎に数値テーブルが必要になる。従って、数値テーブルの数若しくはデータ量が膨大になるとい

う大きな問題が有る。また、話者適応処理や雑音適応処理のために、混合ガウス分布のパラメータ（平均・分散）を修正すると、それに伴う計算量も膨大になり、数値テーブルを修正するにも多大の処理が必要になる。

- 5        上述のように非線形スカラ量子化を採用した場合には数値テーブルの参照に多大の演算処理を要し、線型形スカラ量子化を採用した場合には数値テーブルの参照は効率化されるが膨大な数値テーブルが必要になり、それ故に適応処理には多大な演算時間を要し、何れにおいても、携帯情報端末装置や、電池駆動されるデータ処理システムのように、演算処理能力が比較的低いデータ処理システム、更には低コストの要請が
- 10       厳しいデータ処理システムでは、実用に耐えることができない。

本発明の目的は、HMMの出力確率を高速に計算できると共に、話者適応や環境適応等のモデルの修正に柔軟に対応できるデータ処理システム、そして混合ガウスHMM出力確率演算方法を提供することにある。

- 15       本発明の別の目的は、携帯情報端末装置や、電池駆動されるデータ処理システムのように、演算処理能力が比較的低いデータ処理システム、更には低コストの要請が厳しいデータ処理システムであっても、出力確率演算の高速化と、適応による多次元ガウス分布の変更に対する処理の高速化とを実現できるデータ処理システムを提供することにある。

- 20       本発明の上記並びにその他の目的と新規な特徴は本明細書の以下の記述と添付図面から明らかにされるであろう。

#### 発明の開示

##### 《中間テーブルによる可変マッピング》

- 25       混合ガウスHMMでは、出力確率は、混合多次元ガウス分布（式2）のような関数で与えられる。例えば、混合多次元ガウス分布は、多次元

5      ガウス分布の和になり、多次元ガウス分布は特徴成分毎の1次元ガウス分布の積になる。特徴成分は認識対象音声の観測系である特徴ベクトルの成分である。特徴成分毎の1次元ガウス分布の分散及び平均は特徴成分毎に固有である。種々の1次元ガウス分布の数値をテーブル化するとき、特徴成分毎に個々の1次元ガウス分布の数値テーブルを用意することはしない。中間テーブル(301, 401)を設ける。即ち、数値テーブル(1052)には代表的な分散及び平均を持つ複数種類の1次元ガウス分布を基にした夫々の分布の数値が格納されている。特徴成分に対しては線形スカラ量子化を採用し、その量子化値をインデックスとして中間テーブル上の情報を参照する。中間テーブルを特徴成分毎に設ける場合、個々の中間テーブルは、所要の分散及び平均に応ずる1次元ガウス分布に関する数値テーブル上の数値の所在を示すアドレス情報が格納されている。適応によって1次元ガウス分布の分散や平均を変更する場合、当該変更された分散や平均に応ずる1次元ガウス分布の数値データの所在に従って中間テーブルの内容を書き換える。

10      15

各特徴成分に共通のグローバルテーブル(400)を形成し、このグローバルテーブルから中間テーブルを抽出して用いるようにすることも可能である。グローバルテーブルは、第17図に例示されるように、X-Y方向にマトリクス状に記憶領域を有し、個々のX方向の配列は数値テーブル上の対応する1次元ガウス分布の数値の所在を示すアドレス情報の配列とされ、夫々のX方向の配列に関する1次元ガウス分布の分散は相互に相異され、その平均は、例えば分布の中央に統一されている。グローバルテーブルに対するY方法の選択には1次元ガウス分布の分散の値を考慮し、X方向の先頭位置の選択には1次元ガウス分布の平均の値を考慮する。平均が大きいほどX方向の先頭位置をX方向寄りにシフトすればよい。グローバルテーブルのY方向位置とX方向先頭位置

20      25



とによって当該X方向先頭位置から始まる中間テーブルを抽出できる。抽出された中間テーブルに対するアクセスは、前述と同様に、特徴成分の量子化値を当該先頭位置からのオフセットとして用いる。適応によって1次元ガウス分布の分散だけを変更する場合、中間テーブルを抽出する  
5 ときのY方向位置を変えればよい。適応によって1次元ガウス分布の平均だけを変更する場合、中間テーブルを抽出するときのX方向先頭位置を変更すればよい。特徴成分毎に抽出すべき中間テーブルの先頭アドレスはアクセスポインタ ( $P_0 \sim P_n$ ) によって指示すればよい。アクセスポインタの値は、分散 ( $\sigma$ ) や平均 ( $\mu$ ) に応じて予め演算しておくことができる。適応に際しては、分散や平均の修正に応じてそのアクセスポインタの値を修正しておくことができる。特徴成分毎のアクセス  
10 ポインタは、多次元ガウス分布毎にアクセスポインタテーブル (420) に予め纏めておくことができる。

以上のように、各特徴成分毎に数値テーブル参照のための複雑なパラ  
15 メータ演算を回避しつつ、平均や分散の修正に対応するために、線形スカラ量子化を採用し、また、線形量子化された特徴成分に応じた数値テーブルのアクセスのパターンを制御するために、中間テーブルを採用した。線形量子化された特徴成分と数値テーブルとの間に、マッピング関係を可変とするインデックス変換を目的とした中間テーブルを挿入す  
20 ることで、適応による分散や平均の修正に対して容易に対応できる。即ち、そのような適応による分散や平均の修正に対して、前記グローバルテーブルを用いる構成ではアクセスポインタを修正するだけで対応することができる。見方を変えれば、線形スカラ量子化とインデックス変換を目的とした中間テーブルを組み合わせることで、線形スカラ量子化  
25 による数値テーブル参照の高速化を保証しつつ、非線形スカラ量子化と同じようにテーブルのデータ量削減を実現することができる。

### 《インデックス変換の類型化と共通化による効率化》

上記の構成を単純に実現すると、数値テーブルの書き換えは発生しないが、その代りに中間テーブルの書き換え等が発生する。この問題に対処するために、第1に、(a) インデックス変換の類型化による中間変換

5 パターンを事前に計算しておく構成を採用する。即ち、話者適応化あるいは環境適応化において、適応化はガウス分布の平均および分散の修正変更によって行われる。この平均および分散のパターンを類型化し、これを事前に保有することで、テーブル変更のコストを最小にする。第2

10 に、(b) 中間テーブルの共通化による簡略化を行なう、即ち、上記の方法においては、各々のHMM毎に、各々の混合分布で中間テーブルを有することを想定していたが、これは、全ての変換パターンを網羅したテーブルが一つあれば、そのテーブル上の（各HMMの各混合分布の）アクセス位置を保有することで、中間テーブルの機能が実現できる。この

15 場合、話者適応および環境適応化は、上記のアクセス位置の修正のみで十分である。

### 《中間テーブルによる計算分布の選択》

混合ガウス分布の計算で、計算分布の削減は計算高速化の有効な方法である。本発明では、この中間テーブルに、分布選択機能を具備することで、計算の簡略化を図る。一般に、多次元ガウス分布は、各特徴次元

20 における一次元ガウス分布の積で表現されるが、この各々の一次元ガウス分布に対する評価を中間テーブル中に挿入することで、数値テーブルの無用な参照を減らして、分布の削減機能を実現できる。

### 《データ処理システム》

本発明の一つの態様であるデータ処理システムは、特徴ベクトルに対してHMM音声認識を行なうためにデータプロセッサ(103)が中間

25 テーブル(301, 302)及び数値テーブル(1052)を参照して

混合多次元ガウス分布によって表現される出力確率を演算可能であって、前記数値テーブル（1052）は、複数種類の1次元ガウス分布を基にした夫々の分布の数値を格納する領域（1052E）を有し、前記中間テーブル（301, 302）は、前記特徴ベクトルの特徴成分の値  
5 に対する線形量子化値に基づいて選択される領域にその量子化値に対応する前記数値テーブルの値の所在を示すためのアドレス情報を格納する領域（301E, 302E）を有する。そして、前記データプロセッサは、前記特徴成分の値を線形量子化すると共に、特徴成分毎のアクセスポイント（310のP0～Pn）により中間テーブルを選択し、前  
10 記線形量子化された値を基に前記選択した中間テーブルよりアドレス情報を取得し、取得したアドレス情報を用いて数値テーブルを参照する処理を行い、数値テーブルから参照した値に基づいて前記出力確率を演算する。

上記データ処理システムにおいて、混合多次元ガウス分布の多次元ガウス分布毎に特徴成分毎の前記アクセスポイントが配置されるアクセスポイントテーブル（310）の形成領域を有し、データプロセッサは、  
15 前記アクセスポイントテーブルのアクセスポイントを用いて中間テーブルの選択を行なうように構成することができる。

前記量子化に関しては、前記1次元ガウス分布を基にした夫々の前記分布全体を2のN乗個の数値によって表現すると、前記特徴成分の量子化値はその値の上位Nビットになる。これは、単なる特徴成分のシフト動作だけで量子化できることを意味する。  
20

前記データプロセッサは、前記数値テーブルを参照する処理を特徴成分毎に繰り返して多次元ガウス分布の値を演算し、この多次元ガウス分布の値を演算する処理を所定回数繰り返して混合多次元ガウス分布によって表現される出力確率を演算することができる。  
25

中間テーブルに分布削減のための距離情報を入れておくことができる。前記中間テーブルは、前記数値テーブルの基準になる 1 次元ガウス分布の平均位置を起点に分散の複数倍の範囲に対して前記アドレス情報を格納する領域 (E 1) を有し、その外側には、前記平均からの距離  
5 情報を格納する領域 (E 2) を有し、前記データプロセッサは、多次元ガウス分布の値を演算するために前記数値テーブルを参照する処理を特徴成分毎に繰り返していくとき、中間テーブルから参照した情報が前記距離情報であるときこれを累積し、その累積値が所定値を越えたとき、  
10 当該多次元ガウス分布のための演算を中止するようにすることができる。

別の分布削減情報として、前記中間テーブルには、前記距離情報の外側に固定値 (例えば値 “0”) を格納する領域 (E 3) を設け、前記データプロセッサは、前記中間テーブルから前記固定値を参照したとき、  
15 現在処理中の当該多次元ガウス分布のための演算を中止するようにすることができる。

データ処理システムは例えばバッテリー (1 2 1) を動作電源とする携帯情報端末装置 (1 2 0) 等として構成することができる。バッテリー駆動される装置は低消費電力の要請が厳しく、前述の出力確率の演算  
20 負荷を低減できることから、前記データプロセッサは消費電力が 1 W 以下のものであっても、実用に耐える高速で音声認識処理を行うことができる。

#### 《グローバルテーブルを用いるデータ処理システム》

グローバルテーブルを用いることに特化したデータ処理システムは、特徴ベクトルに対して HMM 音声認識を行なうためにデータプロセッサ (1 0 3) がグローバルテーブル (4 0 0) 及び数値テーブル (1 0  
25 5 2) を参照して混合多次元ガウス分布によって表現される出力確率を

演算可能であって、前記数値テーブル(1052)は、相互に平均が同一であって分散の異なる複数種類の1次元ガウス分布を基にした夫々の分布の数値を格納する領域(1052E)を有し、前記グローバルテーブル(400)は、前記数値テーブルにおける分布毎のX方向の配列をY方向に複数組格納する領域(400E)を有し、前記X方向の配列は、前記特徴ベクトルの特徴成分の値に対する線形量子化値に基づいて選択される位置にその量子化値に対応する前記数値テーブルの値の所在を示すためのアドレス情報を格納する。前記データプロセッサは、前記特徴成分の値を線形量子化すると共に、複数組のX方向の配列に対するY方向の選択に分散が考慮されX方向の配列に対する先頭位置の決定に平均が考慮された特徴成分毎のアクセスポイント(第38図のP0~Pn)の値に従って前記グローバルテーブルから中間テーブル(401, 402)を抽出し、前記抽出した中間テーブルの先頭位置を起点に、前記線形量子化値に基づいて前記アドレス情報を取得し、取得したアドレス情報を用いて数値テーブルを参照する処理を行い、数値テーブルから参照した値に基づいて前記出力確率を演算可能である。

データプロセッサは、アクセスポイントテーブル(420)のアクセスポイント(P0~Pn)を用いて中間テーブルの抽出を行なうことができる。アクセスポイントテーブルは、混合多次元ガウス分布の多次元ガウス分布毎に特徴成分毎の前記アクセスポイントが配置されたテーブルである。

前記データプロセッサは、適応によって混合多次元ガウス分布の平均と分散の双方又は一方が変更されるとき、これに応じて前記アクセスポイントテーブルのアクセスポイント値を変更すればよい。グローバルテーブルそれ自体の内容を修正するに及ばない。

前記アクセスポイントテーブルを複数セット形成しておけば、前記デ

ータプロセッサは、話者を識別し、その識別結果に応じたアクセスポイントテーブルを用いることができる。

前記話者の識別を話者を明示するスイッチ（1302SW）の状態に基づいて行なうことができる。例えば、トランシーバのように片方向通話を行なうデータ処理システムにおいて送話と受話毎の切換えに連動して話者識別を行なうことができる。

前記アクセスポイントテーブルと話者とを対応付ける管理テーブル（500）を採用することができる。このとき、前記データプロセッサは、話者の特徴を示す予め登録した識別用特徴情報と実際の音声特徴分析結果との比較結果に基づいて前記話者の識別を行ない、この識別された話者が、前記管理テーブルに登録されている話者であるときは、当該登録話者のアクセスポイントテーブルを参照する。

前記データプロセッサは、前記管理テーブルに登録可能な話者の人数を一定に限定すると共に、登録話者毎の使用頻度の情報を前記管理テーブルに追加し、音声特徴分析結果が登録話者であるとき、分析結果に一致する登録話者の使用頻度をインクリメントし、分析結果に不一致の登録話者の使用頻度をデクリメントし、音声特徴分析結果が登録話者以外であるときは、最低使用頻度の登録話者を前記管理テーブルから削除し、これに代えて当該登録話者以外の話者を管理テーブルに追加するように構成することも可能である。

複数の音声入力系を持ち、個々の音声入力系毎に、前記アクセスポイントテーブルを有し、前記データプロセッサが、前記複数の音声入力系に対して、独立にアクセスポイントテーブルを用いて、並列的な音声認識を行うことも可能である。

前記データプロセッサは、特徴ベクトルの全ての特徴成分に対して、線形量子化を行なうと共に、前記量子化値と前記X方向の配列の単一配

列要素のアドレス量との積に基づいて、抽出される中間テーブルの先頭位置からの特長オフセットを算出し、その後、多次元混合ガウス分布毎に、前記アクセスポイントと特徴オフセットとにより中間テーブルの参照を行って数値テーブルの参照を行うことができる。これにより、混合

5 多次元ガウス分布毎に特徴オフセットの演算をやり直す必要はない。

前記データ処理システムで実行される音声認識のための出力確率の演算制御プログラムは、コンピュータ読み取り可能な記録媒体を介して、データ処理システムに提供することはできる。

10 図面の簡単な説明

第1図はマイクロコンピュータを用いた音声認識システムの一例を示すブロック図である。

第2図はマイクロコンピュータの一例を示すブロック図である。

第3図は第1図で示された音声認識装置を用いて実行される処理の概要を全体的に示したフローチャートである。

15

第4図は認識処理の概要を示すフローチャートである。

第5図はHMMの一例を示す説明図である。

第6図は left-to-right 型のHMMのモデルの一例を示す説明図である。

第7図は混合多次元ガウス分布の一例として3混合2次元の混合多次元ガウス分布の様子を示す説明図である。

20

第8図は2次元の特徴空間を第7図の断面1で切断し横から見た様子を示す説明図である。

第9図は線形スカラ量子化を行う場合の数値テーブルと1次元正規分布との関係を示す説明図である。

25

第10図は線形スカラ量子化原理を例示した説明図である。

第 1 1 図は 1 次元のガウス分布の平均、分散の一例を示す説明図である。

第 1 2 図は第 1 1 図に対して平均、分散が相異された 1 次元のガウス分布を示す説明図である。

5      第 1 3 図は分布削減のため中間テーブルのデータの構成を概略的に示した説明図である。

第 1 4 図は中間テーブルにおける分布削減様の距離情報の例を示す説明図である。

10      第 1 5 図は単一ガウス分布に対する中間テーブルの分布削減情報の配置の一例を示す説明図である。

第 1 6 図は中間テーブルの値に応じた処理の分岐について例示的に示したフローチャートである。

第 1 7 図はグローバル中間テーブルの一例を示す説明図である。

15      第 1 8 図は出力確率の演算処理の詳細な一例を示すフローチャートである。

第 1 9 図は適応処理において混合ガウス分布の平均と分散を修正する処理の一例を示すフローチャートである。

20      第 2 0 図は第 1 9 図の適応処理で修正されたガウス分布の分散と平均に対して対応する中間テーブルポイントの値を決定する処理手順の一例を全体的に示すフローチャートである。

第 2 1 図は音声認識のためのシステムを適用した携帯情報端末装置の外観の一例を示す説明図である。

第 2 2 図は第 2 1 図に示される携帯情報端末装置の一例を示すブロック図である。

25      第 2 3 図は携帯情報端末装置において 2 個のマイクを用いて雑音適応を行う場合の処理手順の一例を詳細に示したフローチャートである。



第 2 4 図は携帯情報端末装置を用いたトランシーバ型通話における音声認識の処理手順の一例を示すフローチャートである。

第 2 5 図は携帯情報端末装置を用いたセパレート型通話における音声認識の処理手順の一例を示すフローチャートである。

- 5      第 2 6 図は話者適応及び雑音適応を行う音声認識システムにおける音声認識処理の手順の一例を示すフローチャートである。

第 2 7 図は教師なし話者適応を実行し使用頻度によって登録話者を決める音声認識処理手順の一例を示すフローチャートである。

- 10      第 2 8 図は教師なし話者適応を実行し使用頻度によって登録話者を一定人数に保つようにした音声認識処理手順の一例を示すフローチャートである。

第 2 9 図は話者適応のための識別情報のうち話者管理に関する話者管理テーブルの構造の一例を示す説明図である。

- 15      第 3 0 図は頻度情報によって話者管理テーブルの構造を修正及び変更する処理の一例を示すフローチャートである。

第 3 1 図は初期化によって話者管理テーブルに新しく入れ替えられたリストに対する操作の一例を示す説明図である。

第 3 2 図は話者管理テーブルに既に存在するリストに対する操作の一例を示す説明図である。

- 20      第 3 3 図は第 3 1 図及び第 3 2 図の処理の手順を示したフローチャートである。

第 3 4 図は 2 マイク型雑音適応の原理を示す説明図である。

第 3 5 図はトランシーバ型通話における音声認識の原理を示した説明図である。

- 25      第 3 6 図はセパレート型通話における音声認識の原理を示した説明図である。

第37図は雑音適応に応じてテーブル先頭アドレスポインタの値を修正する操作を原理的に示した説明図である。

第38図はHMMパラメータセットに含まれるグローバルテーブルのためのアクセスポインタテーブルの構造の一例を示す説明図である。

5 第39図はHMMパラメータセットに含まれる中間テーブルのためのアクセスポインタテーブルの構造の一例を示す説明図である。

第40図は多次元ガウス分布を用いた確率演算のためのテーブルアクセス手法を纏めて示した説明図である。

10 第41図は中間テーブルのアクセスと数値テーブルのアクセスとの関係を時系列的に示した説明図である。

第42図は浮動小数点演算をサポートするマイクロプロセッサを用いる場合に好適な1次元ガウス分布の数値テーブルの一例を示す説明図である。

15 第43図は整数演算で対応可能な1次元ガウス分布の数値テーブルの一例を示す説明図である。

発明を実施するための最良の形態

#### 《混合ガウスHMMを用いた音声認識の概要》

20 先ず、混合ガウスHMMを用いた音声認識技術の基本的な内容について説明する。

第5図にはHMMの一例を示してある。これにより、HMMは、マルコフ過程（時点 $t+1$ の状態が、時点 $n$ の状態によってのみ与えられる確率過程）で表現される状態遷移モデルであることが理解されるであろう。

25 音声認識では、この状態を一種の確率的な「音源」と看做している。ここで、確率的という意味は、この状態に存在する場合、常にある決ま

った音が生成されるとは限らず、色々な音が生成される確率が与えられている。これを、一般に出力確率と呼ぶ。

音声認識では、言葉と音とを、この状態を半順序関係を与えて接続したモデルで表現する。具体的には、第6図のような left-to-right 型の HMM が用いられることが多い。

例えば、「あい」という単語を left-to-right 型の HMM で表現することを考える。これを仮に「Word 1」とする。そして、状態 S 1 には「あ」、状態 S 2 には「い」を表現させる。

この時、例えば、必ず「あ」が1フレーム（例えば10ms）、「い」が1フレーム（例えば10ms）であれば、 $S 1 \Rightarrow S 2$  の状態遷移で表現できることになる。しかし、実際は、様々な長さの「あ」の後に、様々な長さの「い」が続くことになる。

このような時間的な「ばらつき」を表現するため、自分に対する状態遷移と隣に対する状態遷移を確率的に表現する。こうすると、「あ」が n フレーム継続した後に、「い」が m フレーム継続して終了する発声パターンが、確率的に（各パターンの生成確率という形で）表現できる。この確率が遷移確率（状態遷移確率）であり、第6図の Word 1 において  $a_1(1,1)$  は状態 S 1 が次に同じ状態 S 1 を採る状態遷移確率であり、 $a_1(1,2)$  は状態 S 1 が次に隣の状態 S 2 を採る状態遷移確率である。

また、「あ」という発声についても、老若男女によって、その音響的性質は大きく異なる。そこで、それらの統計的な出現パターンから、「あ」という発声を表現する状態 S 1 における特徴ベクトルの出力パターンを確率的に表現することで、様々な人の発声パターンをモデル化することができる。この確率的な表現が出力確率である。第6図において Word 1 の状態 S 1 における出力確率は  $b_{11}(y)$  で表現され、Word 1 の状態 S 2 における出力確率は  $b_{12}(y)$  で表現されている。

5 以上のように、HMMは「様々な人の言葉の時間的あるいは音響的なばらつきを表現」するため、人の発声過程を確率的にモデル化したものであり、当然その評価も確率的にならざるを得ない。すなわち、ある観測系列（入力音声の分析結果）が与えられた場合、各単語を表現するモデルにおいて、当該観測系列が得られる確率（尤度）を評価し、最も尤度の高いモデル（若しくはそれが意味する単語）を、認識候補として出力する。

10 前述のように、HMM音声認識では、最も尤度の高いモデルを認識候補として出力する。このためには、各々のモデル毎に尤度を計算する必要がある、それには、状態毎に、状態遷移確率と出力確率との積を演算する必要があり、全体として膨大な計算負荷が予想される。そこで、そのような演算には、例えば、ビタビ（Viterbi）の方法と呼ばれる、一種の動的計画法を用いた処理が行なわれる。

15 ビタビの方法は、複数経路存在する状態遷移経路（パス）のうち、最も尤度の高いパス（最適パス）を選択し、その尤度によって評価する。この計算は、式1のように効率的に実行できる。

$$\alpha_{t+1}(i) = \max \{ \alpha_t(i-1) \cdot a_{i-1,i} \cdot b_i(y_{t+1}), \alpha_t(i) \cdot a_{i,i} \cdot b_i(y_{t+1}) \} \cdots \text{式 1}$$

$$\alpha_1(i) = b_1(y_1) \quad (i = 1)$$

$$\alpha_1(i) = 0 \quad (i \neq 1)$$

20 式1において  $a_{j,i}$  は状態  $j$  から状態  $i$  への状態遷移確率である。 $b_i(y_t)$  は状態  $i$  において状態  $y_t$  を出力する出力確率であり、 $y_t$  は特徴ベクトルの対応する特徴次数の値である。 $\alpha_{t(i)}$  は時刻  $t$ 、状態  $i$  における前向き確率である。

25 このように、HMM音声認識では、各フレーム毎に状態遷移経路の全ての状態の出力確率の値が必要になる。多くの場合、この出力確率は、混合多次元ガウス分布によって与えられる。これを本明細書では、混合ガウスHMMと呼ぶ。

混合ガウスHMMでは、出力確率は、式2の混合多次元ガウス分布のような関数で与えられる。

$$bs(y) = \sum_k \omega_k \prod_i \frac{1}{\sqrt{(2\pi\sigma_{ski})}} \exp\left\{-\frac{(y_i - \mu_{ski})^2}{\sigma_{ski}^2}\right\} \cdots \text{式2}$$

5 混合多次元ガウス分布を表す式2において、例えば3混合2次元の混合多次元ガウス分布は、第7図のように表すことができる。第7図の3混合2次元ガウス分布は、式3のように表現される。

$$\begin{aligned} bs(y) = & \omega_1 \left[ \frac{1}{\sqrt{(2\sigma_{S11})}} \exp\left\{-\frac{(y1 - \mu_{S11})^2}{\sigma_{S11}^2}\right\} \right. \\ & \times \frac{1}{\sqrt{(2\sigma_{S12})}} \exp\left\{-\frac{(y2 - \mu_{S12})^2}{\sigma_{S12}^2}\right\} \left. \right] \\ & + \omega_2 \left[ \frac{1}{\sqrt{(2\sigma_{S21})}} \exp\left\{-\frac{(y1 - \mu_{S21})^2}{\sigma_{S21}^2}\right\} \right. \\ 10 & \times \frac{1}{\sqrt{(2\sigma_{S22})}} \exp\left\{-\frac{(y2 - \mu_{S21})^2}{\sigma_{S21}^2}\right\} \left. \right] \\ & + \omega_3 \left[ \frac{1}{\sqrt{(2\sigma_{S31})}} \exp\left\{-\frac{(y1 - \mu_{S31})^2}{\sigma_{S31}^2}\right\} \right. \\ & \times \frac{1}{\sqrt{(2\sigma_{S32})}} \exp\left\{-\frac{(y2 - \mu_{S32})^2}{\sigma_{S32}^2}\right\} \left. \right] \cdots \text{式3} \end{aligned}$$

式3で表現される前記3混合2次元ガウス分布を例えば2次元の特徴空間 y 1、y 2 に対して表現した図が第7図である。ここで、\*Aの  
15 山は式3の第1項目、\*Bの山は式3の第2項目、\*Cの山は式3の第3項目によって表現される。この2次元の特徴空間を第7図の断面1で切断し、横から見た様子が第8図に示される。前記式2において、kは混合数、 $\omega_k$ は山の高さ、各次元毎の関数

20  $\frac{1}{\sqrt{(2\sigma_{ski})}} \exp\left\{-\frac{(y_i - \mu_{ski})^2}{\sigma_{ski}^2}\right\}$  は1次元正規分布関数である。これにおいて  $y_i$  は特徴ベクトルの次元毎の特徴成分である。式2において、複数の

山が有るのは、同じ語でも老若男女によって音響的な特徴が相異されることに基づく。

式 2 や式 3 等 に示されるような混合ガウス分布の演算の高速化には、計算する分布を大きく限定する方法と計算の一部をテーブル化する方法とが有効である。また、効率化のため、混合多次元ガウス分布を対数  
5 評価することも多いが、整数処理で行う場合も原理的には全く同じである。例えば式 3 の演算の高速化手法について説明する。

計算を高速にするという観点からは、前述のように、特徴ベクトルを幾つかの標準的なパターンに対応させ（ベクトル量子化）、そのパターン  
10 毎に出力確率を定義する方法が可能である。

いま、第 7 図の混合ガウス分布を例を説明する。この例では、例えば、領域 1 に存在する特徴ベクトルに対しては、式 3 で定義される値は、その第 1 項目の値と殆等しい（すなわち、第 2 項目及び第 3 項目のスコアは殆 0）と看做せる。従って、特徴が領域 1 に存在することさえわかれば、式 3 の出力確率は、その第 1 項目の計算（すなわち分布 \* A の計算）  
15 だけで取得することができる。

上記の処理の場合、特徴空間を部分領域に分割して、その部分領域と計算する分布を対応付けることになるが、特徴ベクトルと部分領域との対応には、ベクトル量子化が用いられることが多い。ベクトル量子化とは、特徴空間上の有限個の代表ベクトルを考え、特徴空間上の任意の点を、その点と最も近い代表ベクトルで近似表現する方法である。例えば、  
20 第 7 図に示されるの特徴空間を a, b, c の 3 点で代表すると、領域 1 の特徴ベクトルは a に対応することになる。

このようなベクトル量子化には、効率的な方法が幾つか提案されているが、基本的に、距離が最小となる代表ベクトルを選択する。例えば、  
25 a, b, c のような代表点から各特徴次数の値までの距離計算を行って、

距離が最小となる代表ベクトルを選択すればよい。このベクトル量子化は、混合多次元ガウス分布をそのまま計算するに比べれば演算量は僅少になるが、それでも計算負荷は小さいとはいえない。

5       また、出力確率の演算の一部をテーブル化して演算速度を高速化することも可能である。この場合も、そのテーブルをベクトル量子化によって構成することができる。しかし、ベクトル量子化して出力確率を対応させると、量子化誤差が大きくなり、認識性能が劣化する。

10       そこで、計算を各々特徴次元における計算に分解して、各特徴次元を標準的なパターンに分割し、各々の計算結果をテーブル化する、スカラ量子化の手法を採用することができる。例えば、式 4

$$\frac{1}{\sqrt{(2\sigma_{ski})}} \exp\left\{-\left(y_i - \mu_{ski}\right)^2 / \sigma_{ski}^2\right\} \quad \cdots \text{式 4}$$

15       で示される単一ガウス分布をテーブル化する。即ち、 $y_i$  の値とこの値に  
      応ずる式 4 の値とを対応させた数値テーブルを設ける。これは、対数系か否かで表現する関数も異なるが、基本的な原理は共通である。この  
20       場合、ベクトル量子化とは異なり、量子化誤差は僅少になる。

      スカラ量子化には前述の通り非線形スカラ量子化、線形スカラ量子化がある。混合ガウス分布のスカラ量子化において、各次元毎の関数は、単一の 1 次元正規分布であり、その分布の特徴は平均と分散が分かれば規定できる。

20       非線形スカラ量子化では、数値テーブルの数を減らすために、代表的な平均と分散の 1 次元ガウス分布に関する数値テーブルを設け、種々の平均と分散に対しては、パラメータ演算を行い、パラメータと特徴成分とから数値テーブルを参照する。しかしながら、この手法は、テーブルアクセスのために必ず各特長成分毎のパラメータ演算を行わなければ  
25       ならず、また、テーブルの参照においても、そのように演算されたパラ

メータを用いるアクセスは、テーブルに対して連続した配列のアクセスになるとは限らないので、テーブルを参照するためのアドレス演算も毎回乗算と加算が必要になる。この技術は前記文献“ON THE USE OF SCALAR QUANTIZATION FOR FAST HMM COMPUTATION”, ICASSP 95, pp.213-216 に

5 記載があり、各特徴成分毎乗算、減算、型変換あるいはシフト演算を要するパラメータ演算を伴うことになり、また、テーブルの参照においても、そのパラメータをインデックスとする配列をアクセスすることになる。この場合、連続した配列のアクセスではないので、機械語（アセンブラ）レベルでは、配列のアドレスの計算に乗算と加算（インデックス

10  $\times$  データ長 + 先頭アドレス）の計算も必要になる。従って、命令レベルでは、乗算が 2 回、加減算が 2 回、と型変換あるいはシフトが 1 回、データのロードが 2 回必要（先頭アドレスと数値データ）になる。

この計算を行わずに、数値テーブルの値を得るには、例えば、一般的な線形量子化を行えば可能である。本明細書では、これを線形スカラ量子化と呼ぶ。

15

第 9 図には線形スカラ量子化を行う場合の数値テーブルと 1 次元正規分布との関係が示されている。線形スカラ量子化の場合、特徴を等間隔に量子化する。量子化し易いように、分布全体を 2 の N 乗個に分割すれば、線形スカラ量子化は、特徴成分の上位 N ビットを抽出ことと同義

20 である。第 10 図にはこの線形スカラ量子化の内容が示されている。

線形スカラ量子化では、量子化の代表点は固定であるから、量子化処理は、各フレーム毎に、換言すれば特徴成分毎に、1 回行えば良い。また、代表点はそのままインデックスに相当するから、数値テーブルにおける先頭アドレスと所望アドレスとの差（以下、オフセット）は、イン

25 デックス  $\times$  データ長になり、その演算も全ての分布で同じであり、1 フレーム毎に 1 回実行すればよい。そして、必要な数値テーブルへのアク



セスは、各数値テーブルの先頭アドレスと、オフセットの和で計算できるから、結局、1回の加算と、2回のロード（先頭アドレスと数値データ）で実行される。

5        混合ガウス型HMMの出力確率の計算（式3）では、単一ガウス分布（含む対数系）に相当する計算量を低減することが重要である。このような各特徴成分毎の計算は、出力確率の計算において最も計算負荷の大きい部分であり、計算数は、全モデル数（認識要素数×left to right で接続された状態数であり、第6図の例では2N個）×混合数×特徴次元の数で表現されるため、些細な計算コストの増大が、全体の計算量の増大に直結する。この部分で、線形スカラ量子化は、テーブルのアクセス以外、全く計算が発生しないため、計算効率の観点からは非常に優れている。

15        しかしながら、線形スカラ量子化では、固定された代表点に対して、各分布毎に数値テーブルが必要になるので、前述の通り、数値テーブルの数若しくはデータ量が膨大になり、また、話者適応処理や雑音適応処理のために、混合ガウス分布のパラメータ（平均・分散）を修正すると、それに伴う計算量も膨大になり、数値テーブルを修正するにも多大の処理が必要になってしまう。

20        以下詳細に説明する本発明の実施例では、混合ガウス分布を用いる出力確率演算において、その一部を1次元正規分布のデータテーブルのアクセスに置き換えて演算速度の高速化を図るものであり、このとき、中間テーブル若しくはグローバルテーブルを採用し、出力確率を高速に計算できる線形スカラ量子化の特徴を備えつつ、テーブルのデータ量を少なくでき、また、話者適応、環境（雑音）適応等に柔軟に対応できるようにするものである。

25        《音声認識システムの概要》

第1図には本発明の一実施例に係る音声認識システムのブロック図が示される。第1図に示される音声認識システムは、特に制限されないが、音声認識ボード101、マイクロフォン107及びモニタ（ディスプレイ）108によって構成されている。前記音声認識ボード101は、  
5 1チップのLSI上に全て実現することも可能である。また、モニタ203は、例えば、音声入力装置などに用いる場合、必ずしも必要ではない。

前記音声認識ボード101は、A/D変換器102、マイクロプロセッサ（MPU）103、ROM（リード・オンリ・メモリ）105、R  
10 AM（ランダム・アクセス・メモリ）106によって構成される。前記モニタ108が付加される場合、さらにビデオインターフェース（VIF）104が必要である。

前記A/D変換器102は、マイクロフォン107より入力されたアナログ音声信号を、デジタル信号に変換する。前記ROM105は読み  
15 だし専用メモリで、本音声認識システムのプログラムや必要なデータ（例えば、辞書やHMMパラメータ）が格納されている。前記RAM106は、読み書き可能メモリであり、マイクロプロセッサ103のワーク領域若しくはテンポラリ領域等に利用される。

第2図には第1図に示されたMPUの詳細な一例が示される。MPU  
20 103は、バスインターフェース118を介して、前記ROM105、RAM106、A/D102及びVIF104に接続される。MPU103の動作プログラムは、命令キャッシュ110を介して命令制御ユニット112へ送られてデコードされる。MPU103はそのデコード結果に基づいて演算制御動作を行う。必要なデータは、データキャッシュ  
25 117を介して、ロードユニット114からレジスタファイル111へ、或いはレジスタファイル111からストアユニット115を介してデ

ータキャッシュメモリ 117 に送られる。レジスタファイル 111 に格納されたデータは、必要に応じて、整数演算であれば整数ユニット 116 で処理され、浮動小数点数であれば浮動小数ユニット 117 で処理され、その処理結果は再びレジスタファイル 111 に戻され、前記ストアユニット 115 を介してメモリに書き込まれる。データアクセスにおいてデータキャッシュ 117 がキャッシュヒットであれば外部メモリのアクセスは行なわれず、データキャッシュ 117 からのリード、或いはデータキャッシュに対するキャッシュフィルが行なわれる。キャッシュミスの場合には外部データメモリのアクセスが行なわれ、更に、必要なエントリが外部データメモリからデータキャッシュ 117 に追加される。命令アクセスにおいて命令キャッシュ 110 がキャッシュヒットであれば外部メモリアccessは行なわれず、命令キャッシュ 110 から命令がフェッチされる。キャッシュミスの場合には外部命令メモリのアクセスが行なわれ、更に、必要なエントリが外部命令メモリから命令キャッシュ 110 に追加される。

第 3 図には、第 1 図で示した音声認識装置を用いて実行される処理の手順として、電源をオンにしてシステムを立ち上げてから、電源をオフにしてシステムを停止するまでの処理の概要が全体的に示されている。

第 3 図においてステップ 201 は処理の開始を示す。これは、具体的には動作電源の投入（電源オン）などによって指示される当該システムの動作開始に相当する。当該システムの動作を開始すると、ステップ 202 によって、必要なデータ 250 を ROM 105 から読み込み、それを RAM 106 若しくはデータキャッシュ 117 に展開する。この場合、滅多にアクセスしないデータや、書き換えないデータで高速な不揮発メモリを用いている場合、RAM 106 等にあえて展開しなくとも、必要な時に直接 ROM 105 をアクセスしてデータを取得すればよい。

ステップ203～205は、ここでは、一種の無限ループであり、それは例えば終了命令を実行するまで繰り返される。ステップ205で終了が判定されると、システム動作が終了される（ステップ206）。この間、適応処理（ステップ203）と認識処理（ステップ204）は必要に応じて実行される。

適応処理とは、必要に応じてHMMなどの諸パラメータを修正する処理を意味する。例えば、環境適応を例に採れば、使用する雑音環境における雑音をサンプルし、それに応じてHMMの出力確率を修正する。出力確率が前記式2で表わされる混合ガウスHMMでは、各混合ガウス分布の平均と分散の修正を意味する。データ252は適応のためのデータ、データ253は認識のためのデータである。

認識処理（ステップ204）は、必要に応じて上記の適応処理（ステップ203）されたHMMパラメータ（データ251）を用いて実行される。ここでは、マイク107からの入力音声データ253に対して音声認識を行い、認識された結果254（例えばテキストデータ）を出力する。

第4図には前記認識処理（ステップ204）の概要が示されている。ステップ211によって認識処理が開始されると、まず、ステップ212で、サンプル音声253の特徴が分析される（特徴分析）。

特徴分析は、音声波形を、一定間隔（例えば10ms毎）で一定区間取りだし（この部分音声区間のことをフレームと呼ぶ）、このフレームにおいて、音の性質が変化しないのものとして（定常性のあるものとして）、音の性質を分析する。音の性質は、例えば周波数スペクトル（FFTによって計算できる）やLPC係数（Levinson-Durbinの再帰式によって計算できる）により分析できる。これらは、一般に複数のパラメータ群によって表現されるため、特徴ベクトルと呼ばれる。この特徴分

析により、音声信号 2 5 3 は、フレーム毎の特徴ベクトル 2 5 5 に置き換えられる。尚、 $n$ 次元の特徴ベクトルは  $n$ 種類の周波数成分を有する。この特徴ベクトル列を観測ベクトル列と呼ぶ。

5 次のステップ 2 1 2 によって、出力確率の計算が行われる。第 5 図に基づいて説明したように、HMMにおいて、出力確率とは、各々の状態が「ある特徴」の音を出力する確率を意味する。従って、出力確率は、前記式 2 で説明したように、「ある特徴」を示す特徴ベクトルの関数として表現される。

10 HMM音声認識には、特徴ベクトルをベクトル量子化し、その量子化ベクトルの関数として出力確率を与える方法（離散型HMM）と、特徴ベクトルの確率関数として与える方法（連続型HMM）があるが、本実施例では、後者のうち、出力確率を混合ガウス分布で定義した方法を対象としている。

15 混合ガウス型HMMの場合、出力確率は、特徴ベクトルの関数として、各HMMの状態毎に、前記式 2 で与えられている。

この出力確率計算は、ステップ 2 1 4 の認識照合時（ビタビサーチ）に並行して行うことも可能であるが、計算負荷が大きいことから、重複した計算を避けるため、照合（サーチ）2 1 4 に先だって、必要な出力確率を計算している（ステップ 2 1 3）。

20 ステップ 2 1 4 では、ステップ 2 1 2 によって得られた観測ベクトル列とステップ 2 1 3 によって計算された出力確率 2 5 6 から、各モデルのスコアを計算する。ここで、スコアとは、例えば第 6 図に示したモデルが、与えられた特徴ベクトル列のパターンを生成する（対数）確率で定義することができる。認識候補は、スコアの最大となるモデルとする。

25 各モデルにおいて最も確率が高くなる状態遷移系列のスコア（以下、ビタビスコア）を当該モデルのスコアと看做して、ビタビサーチが行なわ

れる。

《中間テーブルを用いた出力確率の計算》

第18図には本実施例における出力確率の演算処理（ステップ213）の更に詳細が示される。

- 5       本発明では、単一ガウス分布の（対数）確率計算を、特徴成分を均等に部分領域分割し（線形スカラ量子化）これに対応した計算結果を、予め数値テーブル化し、計算負荷を低減している。線形スカラ量子化を行う利点は、各特徴に対して、全ての混合分布について同一の点に量子化される。すなわち、量子化処理が各分布全てに共有されるため、1フレームに1度で済む。また、数値テーブルのインデックスが、各特徴成分で共通化されると、数値テーブルのオフセット（アクセスすべきテーブルの先頭アドレスと該当配列要素のアドレスの差＝一般にインデックスとデータ長の積で計算される）も同一となるため、数値テーブルのオフセットを求める処理も1フレームに1度で済む。これによって、（非線形スカラ量子化と異なり）単一ガウス分布の計算に必要な処理が、加算（配列の先頭アドレスとオフセットの和）とロードストアだけで可能になり、非線形量子化に比べて著しく短い計算時間で計算できる。
- 10
- 15

- しかし、このようなアプローチでは、適応などの処理によって分散、平均が修正されると、（特徴の対応関係は固定であるから）数値テーブルの変更が必要になる。この変更を避けるため、数値テーブルへのアクセスアドレスが設定された中間的なテーブル（中間テーブル）を用いて、アクセスパターンを制御する。また、この中間テーブルに分布の選択、削減のための情報を持たせて、計算を簡略化する。以下、その内容を詳述する。
- 20

- 25       ステップ1000は出力確率計算の前記ステップ213の開始を意味する。ステップ1001では、ステップ212で分析された特徴ベク

トル（整数型・浮動小数型いずれでも可能）に対し、線形スカラ量子化を行い、その値（インデックス）対して、オフセット（以下、特徴オフセット若しくはテーブルオフセットと称する）を計算する。この計算は

5    された値を全体の量子化数で割り、それにデータ長（一つの配列全体のデータ長）を掛ける処理によって、特徴オフセットを計算できる。線形量子化は第 10 図で説明したように、量子化範囲を 2 の N 乗個に分割すれば特徴成分の上位 N ビットを得ることによって量子化できるのであるから、量子化数／データ長の値を 2 の N 乗の形式にすれば、1 回の右

10   シフト）で実行できる。式で示せば、浮動小数系の場合、特徴成分にある定数（定義域長／量子化数×データ長）を掛けて整数型に変換する。

これ以降、第 18 図の処理では、前記特徴オフセットを使用し、特徴ベクトルは計算には用いない。この特徴オフセットは、データ 1050 で表現する。

15    ステップ 1002 では、ステップ 1001 で求めた特徴オフセットから、各状態の各分布毎にアクセスすべき中間テーブルのアクセスアドレスを求める。中間テーブルのアクセスアドレスは、各分布ごとに定義した中間テーブルの先頭アドレス（当然、各分布ごとに全て異なる）と特徴オフセット（同じ特徴次元に対して、全て共通）を加えて求められる。

20    中間テーブルは、第 11 図及び第 12 図に例示される 301、301、302 の 1 次元ガウス分布に 1 対 1 対応で配置される形態、或いは、第 17 図に例示される 401、402 のように複数の特徴成分に兼用可能なグローバルテーブル 400 から抽出する形態の何によっても構成することができる。後者において、グローバルテーブル 400 は多数の中間

25   間テーブルの集合として位置付けることができる。第 11 図及び第 12 図において 301、302 で示されるものが中間テーブルの一例である。

第17図において400がグローバルテーブルの一例を示す。第17図において401、402はグローバルテーブル400から抽出された中間テーブルの一例である。

例えば第17図に示したグローバルテーブル形式において、中間テーブル401、402の先頭アドレスは、グローバルテーブル400から  
5 中間テーブルとして抽出すべきデータ領域の先頭位置、例えばP1、P2を示している。この先頭位置の決定手法について詳細は後述するが、第38図に例示されるように、特徴成分毎の平均及び分散の値が格納されたテーブル410の値を用いて演算し、或いはその演算結果を予め蓄  
10 えたアクセスポインタテーブル420を用いることができる。アクセスポインタテーブル420のポインタP0～Pnが特徴成分毎に抽出すべき中間テーブル401、402の先頭位置を指している。

一方、第11図及び第12図に例示した中間テーブル301、302の形式では、中間テーブル301、302の先頭アドレスは個々の中間  
15 テーブル301、302の先頭アドレスを意味する。特徴成分毎に定義されるべき中間テーブルの先頭アドレスは、例えば、

第39図に例示されるように、特徴成分毎に、アクセスポインタP0～Pnとしてアクセスポインタテーブル310に定義することができる。

前記アクセスポインタテーブル310、420などは第18図においてインデックステーブル1051と称されている。尚、第18図においてテーブルアドレス1055はステップ1002で演算された中間テ  
20 ーブル先頭アドレスに前記特徴オフセットを加算した値である。

この例では、中間テーブル301、401には、数値テーブルのアドレス（オフセット）及び分布削減情報が格納されている。正規分布の場合、第13図に例示されるように、分布の平均（中央値）から一定以上  
25 離れると、数値的には0（対数系で $-\infty$ ）となる。無相関多次元分布は、



一次元正規分布の積で表現されるから、一つの分布でも、中央から非常に離れてしまうと、数値的に計算する意味がない。従って、そのような数値データ不要領域では、中間テーブルのそれに対応する領域には数値テーブルのアドレスを格納せず、例えば、式 6

5      
$$d = -| (y_i - \mu) | / \sigma \cdots \text{式 6}$$

で定義される距離データを格納しておく。式 6 の距離データは、常に負の値になる。更にその外側には値 “0” を格納しておく。特徴成分に対する量子化数が少ない場合には、第 14 図に例示されるように値 “0” を格納しない態様を採用することも可能である。

10      上記距離データと値 “0” が分布削減情報の一例になる。第 15 図には単一ガウス分布に対する前記分布削減情報の配置の一例が示されている。第 15 図において領域 E 1 は数値テーブルのデータのマッピングアドレスが格納された領域、E 2 は前記距離情報が格納された領域、E 3 は前記値 “0” が格納された領域である。平均や分散の値に応じた 1  
15      次元ガウス分布の分布状態によっては E 2, E 3 の領域が無い場合もあるのは当然である。

第 16 図に示されるように、分布削減情報に対しては分布削減条件 1, 2 が判定される。前記中間テーブル 301, 401 からアクセスされた中間テーブルの値を判定し、“0” の場合、当該多次元ガウス分布の値  
20      を “0” とみなし、当該多次元ガウス分布に関する出力確率演算を中断して次の多次元ガウス分布に関する処理に移る。即ち、中間テーブル 301, 401 の値に対して “0” か否かの判定が分布削減条件 1 の判定になる。アクセスされた中間テーブル 301, 401 の値が負数であるときは、その中間テーブル 301, 401 の値を距離情報とみなし、当  
25      該多次元分布中の他の成分の距離情報と積算し、これが一定値を越えていれば、当該多次元ガウス分布に関する出力確率演算を中断して次の多

次元ガウス分布に関する処理に移る。この距離情報の累積値が一定値を越えるか否かの判定が分布削減条件 2 の判定である。中間テーブル 3 0 1, 4 0 1 の値が正数であるときに始めて中間テーブル 3 0 1, 4 0 1 の値を数値テーブルのアドレスとみなし、そのアドレスのデータをフェッチする。

第 1 8 図において分布削減条件 1 の判定 (ステップ 1 0 0 3) では、アクセスされた中間テーブル 3 0 1, 4 0 1 の値が判定され、“0”であると判断されたときは、処理中の当該多次元ガウス分布に関する出力確率演算を中断して次の多次元ガウス分布に関する処理に移る (ステップ 1 0 1 1)。アクセスされた中間テーブル 3 0 1, 4 0 1 の値が負数であるときは、その中間テーブル 3 0 1, 4 0 1 の値を距離情報とみなし、当該多次元分布中の他の成分の距離情報に累計する (ステップ 1 0 0 4)。1 0 5 6 は累計されたメモリ上のデータを意味する。また、アクセスされた中間テーブル 3 0 1, 4 0 1 の値が正数の場合、また、ステップ 1 0 0 4 の累計計算が終了された後、分布削減条件 2 の判定として、距離の累計値が所定値 a を越えているか否かの判定が行われる (1 0 0 5)。これが、一定値を越えていれば、当該多次元ガウス分布に関する出力確率演算を中断して次の多次元ガウス分布に関する処理に移る (1 0 1 1)。

中間テーブル 3 0 1, 4 0 1 の値が正数であるときに始めて中間テーブル 3 0 1, 4 0 1 の値を数値テーブルのアドレスとみなした処理が行なわれる。例えば、ステップ 1 0 0 6 では、第 2 図に示されるようにデータキャッシュ 1 1 7 のようなキャッシュメモリが備えられている場合に、当該アドレスのデータがキャッシュに存在しない場合、R A M 1 0 6 のような外部メモリ上の数値テーブル 1 0 5 2 から前記中間テーブル 3 0 1, 4 0 1 の値で指定されるデータをキャッシュメモリ 1 1 7

に先読み（プリフェッチ）させる。このようなデータプリフェッチはデータバスが空いているとき、適宜行なわれる。これにより、後で、数値テーブルの値を用いて数値累積するとき、必要なデータ 1 0 5 3 は全て若しくは殆どデータキャッシュメモリ 1 1 7 に格納されていることになる。ステップ 1 0 0 7 では、処理中の多次元ガウス分布に関する残りの単一ガウス成分が有るかを判定し、ある場合には、その単一ガウス分布に関する中間テーブルのアクセスアドレス計算（加算演算）に戻って（ステップ 1 0 0 2）同様の処理を行う。このとき、テーブルオフセットの計算はやり直す必要はない。前述の通り、特徴ベクトルの特徴成分は既に線形スカラ量子化されているからである。

第 1 8 図の処理では、第 1 のループ（ステップ 1 0 0 2 ～ステップ 1 0 0 7）で、全ての特徴に対して中間テーブル 3 0 1，4 0 1 をアクセスする。こうすると、分布削減時に無駄となる計算が少なくでき、かつプリフェッチ（中間テーブルの値を用いた数値テーブルのデータプリフェッチ）による遅れも生じない。例えば、一つの多次元ガウス分布に関する処理を行っているとき、分布削減条件 1 の判定で、数値“0”を判定したとき、当該一つの多次元ガウス分布に関する処理を中断できるから、そのような状況に至る場合であっても、無駄に成る処理は最小限に抑えられている。

ステップ 1 0 0 7 の分岐処理を行わず、ステップ 1 0 0 6 の後に、すぐにステップ 1 0 0 8 を行うことも原理的には可能である。しかし、この場合、プリフェッチが有効に機能しない（一般に、メモリからキャッシュにデータ転送するには多少時間がかかる）。また、分布の削減時にも数値テーブルへのアクセスが発生し、好ましくない。

従って、本実施例では、ステップ 1 0 0 8 で、計算が必要な分布に限り、数値テーブルへアクセスして単一ガウスの（対数）値を求める。こ

のとき、数値データは、常にキャッシュメモリに存在し、キャッシュミスミスペナルティは生じない。

多次元ガウスの（対数）値は、この単一ガウスの（対数）値からを計算する。この計算は、全ての単一ガウス分布の値の積（対数値の場合は和）で計算される。従って、ステップ1008では、単にテーブル値を得るだけでなく、これを、既に累積されている値（データ1057）に掛け合わせる（対数系では足し合わせる）処理を行う。この場合、最初の成分を計算する時には、累積の初期値として1（対数系では0）が初期値として必要になる。その累積値は1057として図示されている。

第2のループ（ステップ1008～ステップ1009）で、全ての成分に対してステップ1008の処理が実行されると、その累積結果は、多次元ガウス分布の値となる。したがって、ステップ1010では、原理的には、レジスタに格納されている累積値をメモリにセーブする処理となる。更に未処理の多次元ガウス分布が有れば（ステップ1011）上記処理1002に戻る。上述と同様に、テーブルオフセットの計算は新たに行う必要はない。

もっとも、この多次元ガウスの値は、複数の分布の値を混合しなければならない。混合は全ての値の和（対数系では  $\text{ADDLOG} \rightarrow \text{addlog}(a,b)=\log\{\exp(a),\exp(b)\}$ ）によって混合されるので、同演算を累積値と実行し、それを新しい累積値としてレジスタに格納する（ステップ1010）。

この累積値1058を先の累積値1057と区別するため、これ以降、データ1057の累積値を多次元累積データ、データ1058の累積値を混合累積データと呼ぶ。全ての単一多次元ガウス分布について混合累積データ1058が計算されると、ステップ1012によって、出力確率256が計算される。基本的には、この混合累積が出力確率256と

なるが、数式処理の仕方によっては、必要な定数データ 1 0 5 4 を付加することもある（対数系の処理でパラメータを分離するなどして数値テーブルを少なくするなど）。この場合、定数テーブル 1 0 5 4 から必要なデータを取りだし、値を整えることもある。そして最終的に出力確率  
5 2 5 6 が計算される。

この第 1 8 図に示した処理によって、一つの混合ガウスの計算が処理されたことになる。この処理は、計算すべき混合分布の全てに実行される（一般的な CMHMM の場合、全ての HMM の状態に対して、出力確率が定義されており、その場合、これら全てに値を求めなくてはならない）。したがって、第 1 8 図による計算の簡略化の効果は、これら全ての確率計算に及ぶことになる。

第 1 9 図及び第 2 0 図には、第 3 図のステップ 2 0 3 の適応処理の一例が示される。第 1 9 図では、いわゆる環境適応と呼ばれる適応処理において、HMM パラメーター、具体的には混合ガウス分布の平均と分散  
15 を修正する例が示してある。第 2 0 図には、修正したガウス分布の分散と平均から、各 1 次元ガウス分布毎の中間テーブルのポイントを決定し、更新する処理手順が示されている。

第 1 9 図に示される処理を詳細に説明する。ステップ 1 1 0 1 によって処理が開始されると、ステップ 1 1 0 2 で雑音データの特徴を分析する。これは、例えば、周波数スペクトルを用いるのであれば、FFT (Fast  
20 Fourier Transform: 高速フーリエ変換) などで実行できる。ステップ 1 1 0 3 では、この分析データを基準に適応の可否を判定する。これは、パラメータを決めた（修正した）時の雑音の性質と現在の雑音の性質との比較によって評価する。

25 比較の基準は、例えば、特徴ベクトルの位相を比較の基準にするとか、周波数スペクトルの相互相関性を評価するとか、様々なアプローチが考

えられる。相互相関性を用いた場合、現在の雑音スペクトル（データ 1 1 5 0）とパラメータ決定時のスペクトル（データ 1 1 5 1）の相互相関を求め、評価値 1 1 5 2 とする。この相互相関は式 7 として例示することができる。

$$5 \quad \max_k = \left\{ \left( \sum_{i=1}^N Si \right) \cdot Ni - k \right\} / (Ps \cdot Pn) \quad \dots \text{式 7}$$

但し、 $Ps = \sqrt{\left\{ \left( \sum_{i=1}^N Si \right) / N \right\}}$ 、 $Pn = \sqrt{\left\{ \left( \sum_{i=1}^N Si \right) / N \right\}}$ である。Nは相互相関性を評価するための学習データの数である。

第 1 9 図では、雑音の特性変動に着目した例を示したが、一定間隔で強制的に適応する方法もある。この場合、ステップ 1 1 0 2 は不要で、  
10 評価値 1 1 5 2 には、時間情報（更新してからの時間）を格納し、一定時間以上経過で適応処理実行と判定すればよい。

いずれの場合でも、適応の判断は、評価値 1 1 5 2 によって判断する。

適応が必要と判断されると、ステップ 1 1 0 5 ～ 1 1 0 7 の処理が行われる。例えば、 $\tau = 1, 2, 3, \dots$ とすると、雑音の特徴ベクトル  
15 を、

$n(\tau) = \{ n_1(\tau), n_2(\tau), \dots \}$ とすると、ステップ 1 1 0 5 によって、雑音データから、例えば式 8 で示されるように平均を修正する。

$$\mu_{ki} = \frac{\sum_{\tau=1}^T K_k(\tau) \cdot n_i(\tau)}{\sum_{\tau=1}^T K_k(\tau)} \quad \dots \text{式 8}$$

但し、 $K_k(\tau) = \frac{\omega_k \prod \left[ \left\{ (1/2\pi\sigma_i)^{1/2} \right\} \exp \left\{ (n_i - \mu_i)^2 / \sigma_i^2 \right\} \right]}{\sum \omega_s \prod \left[ \left\{ (1/2\pi\sigma_i)^{1/2} \right\} \exp \left\{ (n_i - \mu_i)^2 / \sigma_i^2 \right\} \right]}$ 、n は学習デー

20 タである。

同様に、ステップ 1 1 0 6 によって、例えば式 9 で示されるように分散を修正する。

$$\sigma_k = \frac{\sum_{\tau=1}^T K_k(\tau) \cdot n i^2(\tau)}{\sum_{\tau=1}^T K_k(\tau)} - \mu k i^2 \quad \cdots \text{式 9}$$

$$\text{但し、} K_k(\tau) = \frac{\omega k \prod \left[ \left\{ (1/2\pi\sigma i)^{1/2} \right\} \exp \left\{ (n i - \mu i)^2 / \sigma i^2 \right\} \right]}{\sum \omega s \prod \left[ \left\{ (1/2\pi\sigma i)^{1/2} \right\} \exp \left\{ (n i - \mu i)^2 / \sigma i^2 \right\} \right]}$$

- 5      また、ステップ 1 1 0 7 によって、例えば式 1 0 で示されるように混合重みを修正する。

$$\omega k = \frac{1}{T} \sum_{\tau=1}^T k(\tau) \quad \cdots \text{式 1 0}$$

$$\text{但し、} K_k(\tau) = \frac{\omega k \prod \left[ \left\{ (1/2\pi\sigma i)^{1/2} \right\} \exp \left\{ (n i - \mu i)^2 / \sigma i^2 \right\} \right]}{\sum \omega s \prod \left[ \left\{ (1/2\pi\sigma i)^{1/2} \right\} \exp \left\{ (n i - \mu i)^2 / \sigma i^2 \right\} \right]}$$

- 10      ステップ 1 1 0 2 で用いる分析は、必ずしも音声認識で用いる特徴分析法である必要はない。しかし、ステップ 1 1 0 5 ～ 1 1 0 7 での特徴は、音声認識で用いる特徴分析パラメータであることは当然である。従って、仮にステップ 1 1 0 2 で音声認識で用いる特徴分析でなければ（例えば、音声認識が L P C ケプストラムでステップ 1 1 0 2 が周波数スペクトルなどの場合）、ステップ 1 1 0 5 ～ ステップ 1 1 0 7 に先だ
- 15      って必要な処理を実行する。

ステップ 1 1 0 5 ～ ステップ 1 1 0 7 の処理は、全ての混合分布について行われる（ステップ 1 1 0 8）。そして、全ての混合分布について修正した後、本雑音の分析データ 1 1 5 0 を、想定特性 1 1 5 1 に格納し（ステップ 1 1 0 9）、ステップ 1 1 1 0 で終了する。

- 20      《グローバル中間テーブル》

第 19 図の処理によって、混合分布を構成する 1 次元のガウス分布の平均・分散は修正される。この様子は、例えば第 11 図及び第 12 図に例示される。このように 1 次元のガウス分布の平均・分散が修正される時、第 9 図及び第 10 図に示されるような線形スカラ量子化を行ったま  
5 ま、数値テーブルを書き換えることなく、数値テーブルを適切にアクセスできるように、第 11 図及び第 12 図に例示されるような中間テーブル 301, 302 のアクセスの仕方が変更される。

前記中間テーブル 301 を挿入することで、テーブルアクセスが余分に発生するが、第 18 図の処理で説明したように、中間テーブル 301  
10 に数値テーブルのアドレスを格納し、ループ分割・プリフェッチを行えば、数値テーブルの前に中間テーブル 301 が挿入されても、中間テーブル 301 のアクセスによる処理増加は僅少に抑制できる。このことは、第 18 図に基づいて先に示した通りである。

ここで着目すべきは、第 19 図の処理によって分散や平均が修正されたとき、それを中間テーブルにどのように反映するかである。例えば、  
15 中間テーブルに格納される数値テーブルのアドレスを書きかえるならば、分散及び平均の変化に応じて、第 11 図から第 12 図のアクセスができるように該当中間テーブル 301 の内容を中間テーブル 302 の内容に書き換えることができる。第 11 図から第 12 図の書換えを行う  
20 ということは、第 11 図及び第 12 図に示した中間テーブル 301 は、原理的には、全ての 1 次元ガウス分布に対して定義されなければならないということである。しかし、各 1 次元ガウス分布毎に中間テーブル 301 を保有すれば、それだけで膨大なデータ量になってしまうし、平均・分散の修正に伴うテーブル更新コストも同様に膨大になる。

25 ここでは、そのような問題を避けるため、第 17 図に示すグローバルテーブル（グローバル中間テーブルとも称する）400 を一つだけ保有



する。同図にはグローバル中間テーブル400の基本構造を示す。第17図において、白い配列要素は数値テーブルのアドレス（正の値）、黒い配列要素は距離情報（負の値）が格納され、その他は値“0”が格納されている。X方向の配列のデータ領域の数は特徴成分の量子化数よりも大きくされている。これは、1次元ガウス分布の平均の値に応じて中間テーブルの先頭位置がX方向にずらされるため、X方向にはデータ領域を余計に採る必要があるからである。

このグローバル中間テーブル400は、平均（ $\mu$ ）が標準テーブルの平均（ $\mu_0$ ）の場合、様々な分散における数値テーブルのアドレス（オフセット）と前記距離情報が格納されている。第17図の例は、左側の列が分散が最も大きい場合のパターンであり、右にいくほど分散は小さくなる。

このようなグローバル中間テーブル400を作成すると、与えられた平均、分散に対応した中間テーブルのパターンを、グローバルテーブル400上に必ず出現させることができる。即ち、グローバル中間テーブル400の横方向（Y方向）の位置は目的とする1次元ガウス分布の分散（ $\sigma$ ）によって決定する。この分散によって選択されたカラムの配列は、中央が平均（ $\mu$ ）とされる1次元ガウス分布を実現する数値データをアクセスするためのアドレスデータの配列になっている。所望の平均（ $\mu$ ）に対しては、分散（ $\sigma$ ）によって決定されたカラムの配列データに対するアクセス開始位置を平均に応じて縦方向（X方向）にシフトさせて対応する。換言すれば、分散（ $\sigma$ ）によって決定されたカラムの配列データを縦方向にシフトさせる。

例えば第17図において、分布1に対応する中間テーブル401のパターンは、分散が $\sigma$ 、平均が $\mu_0$ の場合、第17図においてP1を先頭アドレスとする配列要素で表現される。同様に、分散が $\sigma'$ 、平均が $\mu$

‘の分布 2 は、第 17 図において P 2 を先頭アドレスとする配列要素から成る中間テーブル 402 で表現される。分布に応じた中間テーブル 401, 402 の先頭アドレス（以下単にアクセスポインタとも称する）P 1, P 2 は、第 38 図に例示されたポインタテーブル 420 として予めテーブル化されていてもよい。前記ポインタテーブル 420 は HMM データの一部を構成する。第 18 図の処理において、特徴成分に対するアドレス計算 1002 では、処理する特徴成分の順番は予め決定しておくことができるので、その順番に従って必要な 1 次ガウス分布を特定できるように中間テーブル先頭アドレスを予めテーブル化して用意しておけばよい。このテーブルは例えば第 38 図のポインタテーブル 420 である。第 18 図のステップ 1001 で演算された特徴オフセットに加算する前記中間テーブル先頭アドレスを、そのテーブル 420 から取り出すことにより、必要な中間テーブルをグローバルテーブル 400 から抽出することができる。

ポインタテーブル 420 を用いることにより、グローバル中間テーブル 400 を、参照専用のテーブル（テーブルの内容を一切書き換えない）とすることができるため、グローバル中間テーブル 400 を他のガウス分布の処理と重複して利用しても（共有化しても）、全く問題を生じない。そして、ポインタテーブル 420 上で定義されたアクセスポインタ（分布 1 では P 1, 分布 2 では P 2）を中間テーブルの先頭アドレスと看做すことで、あたかも実体として中間テーブルが存在するがごとく処理できる。グローバル中間テーブル 400 を用いても第 18 図の処理は全く変わらない。

第 19 図の平均、分散の修正に対応するには、中間テーブルそれ自体の書き換えなど全く必要ではなく、単に、平均、分散に対応するアクセスポインタを計算し、これをアクセスポインタテーブル 420 に反映す

るだけで十分である。すなわち、適応処理によって分散と平均が変化された場合には、元の間テーブル先頭アドレス（アクセスポインタの値）を、分散と平均の変化に応じて変更することにより、中間テーブルの書き換えを行わずに対処できる。例えば、適応前に相当する分布 1 に対応する中間テーブルのパターンが、第 17 図において P 1 を先頭アドレスとする配列要素で表現されているとする。このとき、適応後に相当する分布 1 に対応する中間テーブルのパターンが、第 17 図において P 2 を先頭アドレスとする配列要素に変更されるべき場合には、その分布 1 の中間テーブルの先頭アドレスポインタ（アクセスポインタ）を P 1 から P 2 に変化させるだけでよい。その処理は第 38 図に例示されるようなポインタテーブル 420 に対して行えばよい。

そのための処理を概説すれば、まず、修正された分散と最も近い分散の列（分散カラム）を選択し、そして、平均の修正に対しては、標準ガウス分布の平均と修正された平均との差から、列の先頭位置を上下方向に移動する。

抽出すべき中間テーブルの先頭アドレスの修正について更に詳しく説明する。先ず、標準テーブルを用いて、様々な分散、平均を持つ 1 次元ガウス分布をアクセスできるようにすることを目的とする処理を考える。

$f_0(x_0) = \exp \{-(x_0 - \mu_0)/\sigma_0\}$  とするとき、 $x_0 \Rightarrow f_0(x_0)$  の標準テーブルを用いて任意の平均、分散を持つ  $f(x) = \exp \{-(x - \mu)/\sigma\}$  の値を計算する場合について考える。このとき、 $f_0(x_0) = f(x)$  の関係を満たす  $x_0$  を  $x$  を用いて表現する。

$$f_0(x_0) = f(x) \text{ より、}$$

$$\log\{f_0(x_0)\} = \log\{f(x)\}$$

$$\log\{\exp \{-(x_0 - \mu_0)/\sigma_0\}\} = \log\{\exp \{-(x - \mu)/\sigma\}\}$$

$$(x_0 - \mu_0) / \sigma_0 = (x - \mu) / \sigma$$

$$\therefore x_0 = (\sigma_0 / \sigma)(x - \mu) + \mu_0$$

と式の変形を行うことができる。この式は、平均の位置を原点とするときの  $x_0$  の位置の値  $(x_0 - \mu)$  が、同じく平均の位置を原点とするときの  $x$  の位置の値  $(x - \mu)$  と値  $\sigma_0 / \sigma$  とによって決まる値に等しいことを意味する。上式を更に変形すると、

$$x_0 = (\sigma_0 / \sigma)(x - \mu + \mu_0 \sigma / \sigma_0)$$

となる。ここで、 $\alpha = \sigma_0 / \sigma$ 、 $\beta = \mu - \mu_0 \sigma / \sigma_0$  と置と、

$$x_0 = \alpha(x - \beta) \text{ と表現できる。}$$

そこで、次に、 $C(x) = \alpha(x - \beta)$  とするとき ( $\alpha$ 、 $\beta$  は前述と同じ)、任意の平均及び分散を持つ  $C(x)$  の値を簡単なテーブルを用いて得ることを考える。 $C(x)$  は本来 3 次元テーブル ( $x$ 、 $\alpha$ 、 $\beta$ ) として把握すべきであるが、第 37 図に例示されるように、 $x_0 = \alpha \cdot x$  を定義した 2 次元テーブルを想定し、アクセス時に  $x$  方向に  $-\beta$  ずらして  $C(x)$  を得るようにする。適応後の中間テーブルの先頭アドレスは、 $-\beta$  ずらされて最終的に得られるテーブルの先頭位置に基づいて決定される。第 17 図に従えば、その先頭アドレスは、分布 2 の P 2、即ち、対応される中間テーブルポインタの修正された値になる。

第 20 図には第 19 図の適応処理で修正されたガウス分布の分散と平均に対して前記対応するアクセスポインタの値を決定する処理手順の一例が全体的に示されている。処理が開始されると、標準の平均及び分散値 1 2 5 1 と、適応によって得られた新たな平均値 1 1 5 3 及び分散値 1 1 5 4 とを用いて、前記  $\alpha$  値及び  $\beta$  値を計算する (ステップ 1 2 0 2)。そして、前述の通り、 $\alpha$  値に基づいてグローバル中間テーブル 4 0 0 のテーブルライン (カラム) を決定する (ステップ 1 2 0 3)。更に、 $\beta$  値を用いてテーブル先頭位置を決定する (ステップ 1 2 0 4)。

決定されたテーブルライン及びテーブル先頭値からアドレス計算を行う（ステップ1205）。この計算ではテーブル構造を示すデータ（インデックステーブルのヘッダ）1253が参照される。例えば2次元配列のアドレス計算は、テーブルラインの位置をT、先頭位置をS、1ラインのテーブル要素数をE、1要素のデータ長を4バイト、グローバル中間テーブルの先頭アドレスをA0、アドレスをバイトアドレスとすると、

$$A = A0 + 4 \cdot \{ (T - 1) \cdot E + S - 1 \}$$

によって演算される。Aが、適応後のアクセスポインタの値になる。

- 10 前記ステップ1202～ステップ1205までの処理は、全ての分布に対して繰り返えされる。これによって、第18図の処理で用いる中間テーブルの先頭アドレスは、第17図のグローバル中間テーブル400上のアドレスとして対応付けられる。

- 前述の説明から明らかなように、前記グローバル中間テーブル400
- 15 は平均（ $\mu$ ）と分散（ $\sigma$ ）の値に基づいて参照できるが、上記説明では、抽出すべき中間テーブルの先頭アドレスのポインタ（アクセスポインタ）を格納したポインタテーブル420用いる説明としている。その場合には第38図に例示されるように、特徴成分毎にそのアクセスポインタP0～Pnを備えることになる。アクセスポインタの値は前述の説明
- 20 より明らかなように、分散と平均に基づいて算出することができる。したがって、そのアクセスポインタはそれに対応する分布の分散と平均に一意に対応付けすることができる。したがって、第38図に例示されるように特徴成分毎に分散と平均を備えたテーブル410を用意し、これに基づいてその都度、アクセスポインタP0～Pnの値を演算して求め
- 25 てもよい。但し、前記テーブル410を用いる場合には、適応時の処理は少なくなるが、中間テーブル400を参照するための演算処理が増え

る。逆に、アクセスポインタテーブル 420 を用いる構成では、中間テーブルを参照するための演算処理とデータ量を少なくすることができるが、適応時の処理は増える。前記特徴成分毎の平均及び分散、又は特徴成分毎のアクセスポインタは、混合 HMM のための演算に必要な状態遷移確率などと共に HMM データ（第 3 図の 251）としてシステムに保持されている。

第 42 図乃至第 44 図には 1 次元ガウス分布の数値テーブルの例が示される。第 42 図は、前記式 4 に示される値、即ち第 42 図の矩形 R1 で囲まれた式の値を、所要の分散毎に備えている。数値テーブルが保有する値は、 $-4\sigma \sim 4\sigma$  の範囲とされる。これは、分布削減のための第 13 図の中間テーブルの構成に対応させたものである。この数値テーブルのデータ構造は中間テーブルと共通性が有り、中間テーブルが想定する分散に関する数値データを持つことになる。このような数値データを採用する場合、数値テーブルから参照したデータを積算しなければならないから、演算桁数若しくは演算精度と言う観点から、混合 HMM の演算を行うマイクロプロセッサ 103 は第 2 図に例示されるように浮動小数点ユニットを備えていることが望ましい。

第 43 図に示される数値テーブルは、そのデータを整数演算でも利用可能なように、対数を採った値で数値テーブル化している。この場合には、矩形 R2 で囲まれた式の値が、第 42 図の数値テーブルに格納されることになる。更に、矩形 R3 で囲まれた混合重みの対数値もテーブルに保持しなければならない。第 42 図との大きな相違点は、整数演算でも対応できるという点である。

第 40 図及び第 41 図には以上説明した多次元ガウス分布を用いた確率演算のためのテーブルアクセス手法が纏めて図示されている。

第 40 図において、HMM データには、例えば特徴成分毎にアクセス

ポインタの値がポインタテーブル420として格納されている。例えばある特徴成分のアクセスポインタの値はP1である。適応によってこれがP2に変更されている。このときの演算では、P1から一意に決定される分散及び平均と適応によって変更された分散及び平均とに基づいてアクセスポインタの値P2が決定される。特徴抽出によって特徴成分毎に特徴オフセットが演算され、更にこれに加算されるべきアクセスポインタの値P2が読み込まれて、中間テーブルの参照アドレスが演算される。この参照アドレスで中間テーブル400がリードされると、リードされたアドレスによって、その特徴成分に関する所定の分散及び平均に係る1次ガウス分布の値が数値テーブルから読み出される。

前述の説明から明らかなように、音声認識時の混合HMM演算において特徴成分に応じた1次ガウス分布の取得は、複雑なパラメータ演算などを要せず、アクセスポインタの参照と、特徴オフセットとアクセスポインタとの加算という簡単な処理によって行うことができる。また、適応時は、アクセスポインタを修正すれば良く、グローバル中間テーブル400及び数値テーブル1052の値を修正する必要は全くない。

第41図において、出力確率を演算する前に、特徴ベクトルの各特徴成分に対して予め特徴オフセットを求めておく。そして、特徴成分毎にアクセスポインタの値と特徴オフセットによってグローバル中間テーブル400をアクセスして、1次ガウス分布の数値データのアドレスを取得していく。そして、一つの多次元ガウス分布に含まれる1次元ガウス分布の数値データのアドレスを全て取得したところで、そのアドレスによって数値データをアクセスする。このとき、それまでの間に、その数値データアクセス用のアドレスに対するデータプリフェッチを行っていれば、数値テーブルに対するアクセス時にキャッシュミスを生じない。プリフェッチはMPU103がデータアクセスを行っていない

タイミングで適宜行うことができる。したがって、数値テーブルのアクセス前にグローバル中間テーブル400のアクセスを行っても、数値データの取得が遅れることはない。また、グローバル中間テーブル400をマイクロプロセッサ103内蔵の高速RAM106等に格納してあれば、グローバル中間テーブル400のアクセス時間を実質的に無視し得るほど少なくすることも可能である。適応によって分散及び平均を修正する場合には前述の通り、抽出される中間テーブルの先頭を指すアクセスポインタの値を修正すればよい。

#### 《携帯情報端末装置》

第21図には前記音声認識のためのシステムを適用した携帯情報端末装置120の外観の一例が示される。第22図には携帯情報端末装置120のブロック図が示される。同図に示される携帯情報端末装置120は、特に制限されないが、前記音声認識機能と共に、小型コンピュータ装置の機能を有し、更に携帯電話機能を備えている。ケーシングの中央部にはディスプレイ108とキーボード123が配置され、その端にはマイク（マイクロフォン）107、1301と、スピーカ1307、1308が設けられている。

第22図において、MPU103、ROM105、RAM106、VIF104、ディスプレイ108は第1図で説明した音声認識装置に備えられている回路と同じであり、それらは、前記音声認識機能、小型コンピュータ装置の機能、及び携帯電話機能に共通利用される。

第22図において1303で示されるものは、携帯電話ユニット（PHS）である。携帯電話ユニット1303はアンテナ1309を介してその他の携帯電話や一般の有線電話と通話することができる。スピーカ1307、1308はD/A変換器1305、1306を介してMPU103などに接続される。周辺回路1302は赤外線インタフェース



回路や、フラッシュメモリーカードインタフェース等を実現している。

携帯情報端末装置 120 は、特に制限されないが、2 系統のマイク入力を想定している。マイク 1301 は A/D 変換器 1204 を介して MPU 103 又は PHS 1303 に接続可能にされる。前記マイク 107  
5 は A/D 変換器 102 を介して MPU 103 に接続可能にされる。双方のマイク 107, 1301 は音声認識や電話に利用されるが、その利用形態の詳細については後説明する。

携帯情報端末装置 120 は、携帯性重視の観点から動作電源としてバッテリー 121 を用いる。バッテリー 121 による装置の動作時間を延ばせるようにするために、商用電源を常時動作電源として用いるシステムに  
10 比べて低消費電力が厳しく要請される。これに答えるため、MPU 103 には、動作速度（動作クロック周波数）や、MIPS（Million Instruction Per Second）値、或いは消費電力が比較的小さなものが採用される傾向にある。例えば、消費電力が 1 W 程度、動作クロック周波  
15 数が 200 MHz 程度、データ処理能力が 300 MIPS 程度の MPU 103 を採用することができる。

このとき、MPU 103 を用いて前記音声認識処理を行う場合、混合多次元ガウス分布の計算に、線形量子化の手法とグローバル中間テーブルの手法を採用している  
20 ので、音声認識処理における演算処理の高速化、そして適応時におけるパラメータ変更の高速化が実現されており、そのようなデータ処理能力の比較的低い MPU 103 であっても、音声認識のリアルタイム性若しくは迅速性を阻害することなく、実用に耐え得る程度の速度で音声認識を行うことができる。

混合多次元ガウス分布の計算に、線形量子化の手法とグローバル中間  
25 テーブルの手法を採用した音声認識処理プログラムは、例えば ROM 105 に格納されている。この ROM は、コンピュータである MPU 10

3によってによって読み取り可能な記録媒体である。前記ROM105がフラッシュメモリなどの電氣的に書き換え可能な不揮発性メモリである場合に、前記音声認識処理プログラムを外部から当該ROMにロードして実行することも可能である。例えば、周辺回路1302にインタフェースされる図示を省略するCD-ROMドライブ装置から必要な音声認識プログラムをROMに転送することができる。このとき、前記CD-ROMは、前記音声認識処理プログラムを格納した、コンピュータ読み取り可能な記録媒体の一例とされる。

#### 《2マイク型雑音適応》

10 認識すべき音声から雑音成分をキャンセルするために2個のマイクを用いる公知の技術（例えばANC: Adaptive Noise Canceller）が有る。この技術が採用されているとき、2個のマイクを用いて雑音適応を行う場合について説明する。前記マイク107は主マイクとされ、雑音と共に音声を拾うことができる。これに対して他方のマイク1301  
15 は信号成分に比べて相対的にノイズ成分を大きく拾うようにされた雑音専用の副マイクである。例えばこれは、双方のマイク107、1301の指向性や配置を選ぶことによって実現されている。

第34図には2マイク型雑音適応の原理が示される。音声区間では雑音と音声为重畳され、これが主マイク(107)によってサンプルされる。副マイク(1301)は専ら雑音をサンプルし、そのサンプル信号には音声信号成分は殆ど含まれない。主マイク(107)で得た信号に含まれる雑音の特徴と副マイク(1301)で得た雑音の特徴とは当然相異される。そこで、無声音区間において主マイク(107)と副マイク(1301)の特性が評価されている。例えば、主マイク(107)  
20 の特性を $f_m(\omega)$ 、副マイク(1301)の特性を $f_s(\omega)$ とすると、乗法性の歪を考えれば、 $f_m(\omega) = \alpha(\omega) \cdot f_s(\omega)$ と表現することができる。

無声音区間では、主マイク（１０７）と副マイク（１３０１）とからの信号に基づいて上記 $\alpha(\omega)$ を決定することができる。主マイク（１０  
５ ７）からの入力が入所定の閾値を越えた音声区間では、副マイク（１３０  
１）からの入力に対して雑音分析を行い、 $f_s(\omega)$ の計算を行う。そして、  
 $\alpha(\omega) \cdot f_s(\omega)$ によって  $f_m(\omega)$ の特性補正を行う。その後、前記第  
１９図に示される平均、分散、混合重みの修正を行い、更に第２０図で  
説明したようにポイントテーブル４２０のアクセスポイントの値を修  
正する。

第２３図には、前記携帯情報端末装置１２０において２個のマイクを  
１０ 用いて雑音適応を行う場合の処理手順の一例が詳細に示される。

ステップ２０２のシステム立ち上げでROM２５０からシステムデ  
ータを読み込め、ステップ１４０１で主マイクとして利用されるマイク  
１０７に音声が入力されたか否かを判断する。ステップ１４０２で、音  
声が入力されていないと判断すると、ステップ１４０３を経由し、再び  
１５ ステップ１４０１の処理に戻る。これは、一種の無限ループを構成し、  
主マイクに音声が入力されるまで、繰り返される。

ステップ１４０３では、主マイクに利用されるマイク１０７と副マイ  
クとして利用されるマイク１３０１の特性を比較し評価する。これは、  
音声区間で主マイクから雑音の特性を副マイクの雑音から推定するた  
め、予め主マイクと副マイクの特性の違いを補正するためである。  
２０

ステップ１４０２で主マイクに音声が入力されたと判断されると、ス  
テップ１４０４において、副マイク音声データ（データ１４５１）を副  
マイク雑音分析により特徴分析する（ステップ１４０４）。そして、ス  
テップ１４０３で評価した主副マイク特性（１４５２）を用いて、前記  
２５ ステップ１４０４で得られた分析結果を補正する（ステップ１４０５）。  
そして、前記ステップ１４０４による分析の結果に基づいて、ステップ

1 4 0 6で適応を行うか否かを判定する。適応を行う場合には、前記ステップ1 4 0 5で補正された結果を用いて、雑音適応を行う（ステップ1 4 0 7）。ステップ1 4 0 7の処理は、例えば、第1 9図の手法とほぼ同じ手法（第1 9図において相違点は適応可否の判定に係る処理を行わなくてもよいと言う点である）により実現できる。ここで、修正されたHMMパラメータ（混合ガウス分布の平均・分散）のデータ（1 4 5 3）から、中間テーブルの先頭アドレスを指すアクセスポイントのテーブル4 2 0を更新する処理が行われる（ステップ1 4 0 8）。この処理は、例えば第2 0図の手法で更新することができる。ここで更新されたポイントテーブル4 2 0は、その後の、出力確率計算2 1 2やビタビサーチ2 1 4に利用される。

尚、2マイク型の音声認識では前述のANCの手法の他に、1対のステレオマイクを用いて得られる音声情報を信号成分偏重の情報とノイズ成分偏重の情報に分離してから、前記ANCの手法を採用する公知の技術（例えばビームフォーマ）を適用することも可能である。

#### 《トランシーバ型通話における音声認識》

第2 1図及び第2 2図に例示される携帯情報端末装置1 2 0において音声認識対象は携帯電話ユニット1 3 0 3による通話先からの音声（通話先音声）と、端末装置1 2 0のマイク1 0 7からの入力音声（端末側音声）との2種類ある。前記通話先音声に対する音声認識（通話系音声認識）と端末音声に対する音声認識（端末系音声認識）には、第1に、トランシーバ型通話における音声認識が考えられる。即ち第3 5図に例示されるように、音声を通話先音声と端末音声との何れか一方に切り換え可能にして、双方の音声を排他的に認識可能にする。そのような切り換え操作は、端末系からの音声入力と通話系での受話とを切りかえるスイッチ1 3 0 2 SWで行うことができる。第2 2図においてこのスイッ

チ 1 3 0 2 S W は便宜上周辺回路 1 3 0 2 に含まれる回路として図示されている。双方の音声の特徴は相当異なることが予想される。このとき、H M M の数値テーブルを通話先音声用と端末音声用に別々に持つならば、そのデータが膨大になり過ぎ、また、H M M の数値テーブルを共通化する場合には、通話先音声と端末音声とを切換える毎に適応のための膨大な処理が必要になって、リアルタイム処理が全く不可能になることも予想される。そこで、通話先音声と端末音声とで H M M の数値テーブルと、前記グローバル中間テーブルとを共通化し、前記ポインタテーブル 4 2 0 を通話系音声認識と端末系音声認識とに別々に用意する。そして、別々に用意されたポインタテーブルを入力系毎に使い分ける。通話系音声認識の場合にはそれに割当てられたポインタテーブルを用いてグローバル中間テーブルをアクセスし、端末系音声認識の場合にはそれに割当てられたポインタテーブルを用いてグローバル中間テーブルをアクセスする。尚、第 4 0 図において 4 2 0 - 2 は通話系のポインタテーブル、4 2 0 - 1 は端末系のポインタテーブルを意味する。

第 2 4 図には携帯情報端末装置 1 2 0 を用いたトランシーバ型通話における音声認識の処理手順の一例が示される。

ステップ 2 0 1 で処理が開始されると、ステップ 2 0 2 で R O M 2 5 0 からシステムデータを読み込み、処理が開始される。この例では、端末系からの音声と通話系からの音声とを各々独立に入力できる特徴を生かし、ステップ 1 5 0 1 で、当該音声は、通話系からの音声か端末系からの音声かを判定する。例えば、通話系からの受話と端末系からの送話を切換えるスイッチ 1 3 0 2 S W の状態によって判定する。端末系からの送話入力であることが判定されると、ステップ 1 5 0 3 より端末音声データが音声認識対象に取り込まれる。通話系からの受話入力と判定されると、ステップ 1 5 0 4 より携帯電話ユニット 1 3 0 1 からの通話

系受話音声データが音声認識対象として取り込まれる。ステップ1505では、個々の入力から、無音区間を抽出し、雑音の性質を分析する。ステップ1406では、ここで入力された音声の無音区間のデータを用いて適応するか否かを判定する。適応する場合、適応処理のステップ1407で、分散や平均などのHMMパラメータを修正し、それに応じて、ステップ1408でポインタテーブル420のポインタ値を更新する。これ以降は、第23図と全く同じように処理されるので、その詳細な説明は省略する。

#### 《セバレート型通話における音声認識》

10 第21図及び第22図に例示される携帯情報端末装置120を用いた通話系と端末系の夫々の音声認識の手法として、第2に、セバレート型通話における音声認識が考えられる。即ち第36図に例示されるように、通話先音声（受話音声）と端末音声（送話音声）とを混在させて音声認識可能にするものである。この例においては、前記スイッチ1302 SWは不要である。この場合も事情は上記同様であり、通話先音声と  
15 端末音声とでHMMの数値テーブル及び前記グローバル中間テーブルを共通化し、中間テーブルのポインタテーブルを通話系音声認識と端末系音声認識との夫々に用意する。但し、端末系と通話系の音声区間を別々に検出しなければならない。これによって、通話系と端末系の会話  
20 が重なっても対応できるようになる。尚、グローバル中間テーブルを用いず、特徴成分毎に中間テーブルを割当てて場合には、中間テーブルは通話系と端末系とで夫々別々に持たなければならない。

第25図には携帯情報端末装置120を用いたセバレート型通話における音声認識の処理手順の一例が示される。この例では、通話系と  
25 端末系各々の系に適応させ調整したパラメータセットを2個保有するシステムを構成することになる。この場合、数値テーブル1052やグロ

ーバルテーブル 400 は通話系と端末系で同一であり、中間テーブルのアクセスポインタを保有するポインタテーブル 420 を 2 組持てばよいことになる。

第 25 図において、ステップ 201 で処理が開始されると、まず最初に、ステップ 202 において、システムの立ち上げを行う。本システムでは、端末系音声入力と通話系音声入力 that 別になっていることを利用し、各々の系統毎に処理を行う。ステップ 1503 では、端末系から音声を入力する。ここで、適応の必要があれば、ステップ 1505-1 で無音声区間を検出し、ステップ 1407-1 で雑音適応を行う。そして、この適応に応じて、ステップ 1408 で中間テーブルのポインタテーブル 420-1 を更新する。

これと同様の処理は、通話系でも行われる。本装置のように携帯電話ユニット 1303 と一体になった装置であれば、ステップ 1504 により通話系から認識すべき音声信号を入力する。その後は、前記と同様のステップ 1505-2、ステップ 1407-2、ステップ 1408-2 を行う。

ここで注意すべきは、音声入力系と中間テーブルのポインタテーブルとは夫々 2 系統持つ必要があるが、音声認識処理プログラム並びにグローバル中間テーブル等は単一の (同じもの) で足りる。端末系と通話系各々に別々の認識処理をしているわけではないが、各々に別々の認識処理をするのと同等の性能及び機能を得ることができる。

ステップ 1601 では、重なり調整を行う。これは、端末系と通話系の音声 that 重なった (例えば、一緒に話した) 場合に、調整するものである。これは、簡単な例として、各々の入力音声毎に音声区間検出を行い、先に区間を検出した方の終了を待って、後に区間を検出した方の処理を行うことでも実現できる。

このようにして、音声区間の信号（ただし、端末系か通話系かの区別をするための属性データあるいはフラグを有する）が得られたら、ステップ 2 1 2 で特徴分析、ステップ 2 1 3 で出力確率の計算、ステップ 2 1 4 でピタビサーチを行うことで、系統属性付の認識結果（データ 2 5 4 - 2）が得られる。ここで、系統属性とは、端末系か通信系かを区別する属性データを意味する。

上記処理では、複数系統のデータセットが必要となる処理も、中間テーブルのポインタテーブル 4 2 0 だけ系統毎に保有すれば済む。すなわち、この中間テーブルのポインタテーブルだけを 2 系統持ち、グローバル中間テーブル 4 0 0 や数値テーブル 1 0 5 2 などは端末系と通話系で全て共通でよい。

#### 《話者適応をサポートする音声認識》

第 2 6 図には話者適応及び雑音適応を行う音声認識システムにおける音声認識処理の手順の一例が示される。ここでは、時間情報 1 7 5 2 に基づいて一定時間間隔で適応処理が行われるものとする。

先の例と同様に、ステップ 2 0 1 で処理が開始されると、先ず最初に、ステップ 2 0 2 において、システムの立ち上げを行う。システムが立ち上がると、ステップ 1 7 0 1 により、音声データの取り込みが行われる。そして、ステップ 1 7 0 2 において時間情報 1 7 5 2 をインクリメントする。ここで、時間情報は、クロック単位でも良いし、フレーム単位でも良い。適応を行うべきかの判断（ステップ 1 7 0 3 - 1, 1 7 0 3 - 2）では、時間情報 1 7 5 2 が一定値以上であるかを判別し、一定値以上の場合に適応を実行する。適応しない場合、ステップ 2 1 2 へ移行し、音声認識を開始する。

25 雑音適応を行う場合、先ず、ステップ 1 7 0 4 - 1 で雑音データを入力し、それに応じてステップ 1 7 0 5 - 1 でパラメータを修正する。例



- 例えば、2マイク系では、第23図の方法（ステップ1404～ステップ1407）と同じでもよい。そして、ステップ1706-1において、修正された分散、平均1453に応じてグローバル中間テーブルのアクセスポインターテーブル420を修正し、時間情報1752をリセット
- 5      （例えば、0を設定）する。そして、音声認識処理（ステップ212～ステップ214）を行う。

- 話者適応する場合も同様である。雑音適応の場合と同様に、ステップ1703-2の適応判断では、前記時間情報1752が、一定以上になった場合に適応を実行する。ただし、雑音適応の時間間隔と必ずしも同一でなくとも良い。ステップ1704-2では、雑音適応の場合と異なり、音声区間を抽出する。ステップ1705-2では、所謂教師なしの話者適応を行う。この修正に基づいて、前記ポインターテーブル420を更新する。前記教師なしの話者適応とは、予め、適応のための事前学習を行わない話者適応方式である。
- 10

- 15      上記雑音適応及び話者適応は、一定間隔で、いわば割り込みのごとく発生する。適応を行わない場合には、直接にステップ212へ飛び、音声認識を行う。ここから、ステップ214までは、先に示した例と同様である。

- 第27図には教師なし話者適応を実行する音声認識システムの別の例を示す。ここでは、特に頻繁に使用するユーザを登録し、当該話者の音声では、当該話者向けのポインターテーブルに切り替えるシステムの例である。登録話者以外の場合には一般向けのポインターテーブルに切り替える。
- 20

- 先の例と同様に、ステップ201で処理が開始されると、先ず最初に、
- 25      ステップ202において、システムの立ち上げを行う。システムが立ち上がると、ステップ1701により、音声データの取り入れが行われる。

ステップ 1 8 0 1 では、話者識別のための特徴分析（例えば、高周波の成分の分析）を行う。これによって話者識別様の特徴データ 1 8 5 1 が取得される。

5       ステップ 1 8 0 2 では、前記話者識別用特徴データ 1 8 5 1 と識別情報 1 8 5 2 とを用いて話者識別が行なわれる。例えば、予め話者特徴を識別情報 1 8 5 2 として登録しておいて、話者特徴データ 1 8 5 1 に最も近い登録パターンが有るかを識別して話者を判定することができる。この話者識別（ステップ 1 8 0 2）処理で判定可能な話者に対しては夫々の処理系が設けられている。各々の処理系は、処理（プログラム）  
10       は同一であるが、話者毎並びに一般話者に対して、夫々固有のアクセスポインタテーブルなどのパラメータが設けられている。もっとも、適応可否の判断は、各々の話者に応じて（パラメータに応じて）異なることから、第 2 7 図では、適応処理は話者毎に分離して表現されている。

15       ここでは、登録話者の分とデフォルト（一般話者向け標準パターン）の分とのパラメータセットを用いる。例えば、2 人登録していれば、3 系統のパラメータセットが必要になる。各パラメータセットには、少なくとも、ポインタテーブルが含まれている。

20       ステップ 2 1 2 以降は、先に示した例と同様の認識処理が行われる。ただし、使用されるグローバル中間テーブル 4 0 0 のポインタテーブル 4 2 0 は各々の話者毎に与えられている。グローバル中間テーブル 4 0 0 は全ての話者に共通とされる。このようにすれば、各種テーブルを形成するメモリ容量を抑制することができる。尚、グローバル中間テーブルを話者毎に別々に設けることも可能であるが、その場合には、グローバル中間テーブルによるメモリ使用量が膨大になる。

25       第 2 8 図には教師なし話者適応を実行する音声認識システムの更に別の例を示す。第 2 7 図と同様に、特に頻繁に使用するユーザを登録し、

当該話者の音声では、当該話者向けのパラメータセットに切り替えるシステムの例であるが、特にこの例では、全体の登録話者数を一定数に限定して、使用頻度を考慮するシステムとしている。

5 先の例と同様に、ステップ201で処理が開始されると、先ず最初に、  
ステップ202において、システムの立ち上げを行う。システムが立ち  
上がると、ステップ1701により、音声データの取り入れが行われる。  
ステップ1801では、話者識別のための特徴分析（例えば、高周波の  
成分の分析）を行う。分析された話者識別用特徴データ1851より、  
ステップ1802で話者識別が行なわれる。これには、識別情報185  
10 2を用いる。例えば、予め話者特徴を登録しておいて、最も近い登録パ  
ターンを選択するなどにより実現できる。この話者識別1802におい  
て、処理系が選択される。これら各々の処理系では、処理プログラムは  
同一であるが、使用するポインタテーブルが異なる。もともと、適応可  
否の判断は、各々の話者の特徴に応じて異なることから、第28図では、  
15 話者毎に分離して表現している。以上の点は、第27図の場合と全く同  
じである。

特に第28図の例では、ステップ1901において、識別情報の修正  
を行う。ここでは、第27図で用いた情報の他に、各登録話者の使用頻  
度を管理情報としたテーブル（話者管理テーブル）を使用し、登録者数  
20 を一定数に限定する。この処理を行った後は、第27図で説明した手順  
と全く同じ処理が行われる。

前記識別情報修正処理（ステップ1901）の詳細を第29図及び第  
30図を用いて説明する。第29図では、識別情報1852の内、話者  
管理に関する管理テーブル（単に話者管理テーブルとも称する）500  
25 の構造を示す。ここでは、登録話者の欄501に対して、使用頻度の欄  
502と、ポインタテーブル420へのポインタ（データポインタ）の

欄 5 0 3 とを有し、これらの欄のデータは、登録話者毎に使用頻度順にソート可能にされている。このような話者管理テーブル 5 0 0 は、1 系統のデータセットでは不要であるが、複数系統の場合、必要になる。ただし、第 2 5 図や第 2 7 図の例のように構造が固定の場合（ソート不要の場合）には、あえてテーブル化しなくても、単なる参照データとして

5 データポイントなどの情報を備えていればよい。

第 2 8 図の識別情報修正のステップ 1 9 0 1 では、例では、頻度情報によってテーブル構造の修正及び変更を行わなければならない。これについて簡単に説明する。この処理手順は第 3 0 図に示される。ステップ

10 2 0 0 1 が開始されると、まず、ステップ 2 0 0 2 において、識別された話者に該当する話者がリスト（話者管理テーブル 5 0 0）に存在するか否かを判定する。もし、リストに無い場合、ステップ 2 0 0 3 において、最下位の登録話者と今回の話者とを入れ替える。ステップ 2 0 0 3 のリストの入れ替えでは、最下位のデータを消去し、新しい登録話者の

15 I D（これは、話者認識における登録 I D）を登録話者の欄に書き込み、頻度情報を 1 より大きな値（例えば、5）に設定する。データポイントは前者に割当てられているものを引き継ぐが、該当するグローバル中間テーブル 4 0 0 のポイントテーブル 4 2 0 は、標準パターンに相当するものに設定（初期化）する。

20 ステップ 2 0 0 4 では、頻度情報を更新する。これは、話者識別によって選ばれた話者が登録話者である場合に当該登録話者の頻度情報をインクリメントし、該当しない登録話者の頻度情報をデクリメントする。このようにすれば、初期化してからあまり使用されな話者の頻度情報は、初期化頻度値（本例では、5）より小さくなり、初期化した話者より下

25 位になる。すなわち、初期化し登録されたばかりの話者がすぐにリストから削除されないようにすることができる。

ステップ 2 0 0 5 では、上記操作に伴う順位の変更に対して、使用頻度でソートを行う。ソートの方法は色々あるが、例えば、ディクリメントされたグループの順序関係は保たれているから、第 3 3 図に基づいて後で説明するようなバブルソートで効率的に実行できる。すなわち、初期化されたリストとインクリメントされたリストのみをバブルソートで処理すれば良い。この様子を第 3 1 図乃至第 3 3 図に示す。

第 3 1 図は、初期化で新しく入れ替えられたリストに対する操作の例を示す。この場合、最下位から順にバブルソートする。第 3 2 図は、既に存在するリストに対する操作の例である。この場合、リストが存在する位置からバブルソートする。着目したリスト以外のリストの頻度情報は 1 ずつ減少するため、着目リストの順位は、必ず上昇する方向に動く。従って、着目リスト以外のリストの操作は必要ない。

この手順をフローにしたものが、第 3 3 図である。ここでは、ソートの処理を示す。ステップ 2 1 0 1 で処理が開始されると、ステップ 2 1 0 2 でソートリストが選択される。これは、着目している話者のリストである。ステップ 2 1 0 3 では、直上の頻度情報と比較する。順序関係が正しい場合、ステップ 2 1 0 5 で終了する。順序関係が正しくない場合、直上のリストとリストを入れ替え、ステップ 2 1 0 3 に戻る。この処理は、順序関係が正常になるまで（直上のリストの頻度情報より小さくなるか、最上位に到達するまで）繰り返され、ステップ 2 1 0 5 で処理を終了する。

上記実施例によれば以下の作用効果を得ることができる。

上記出力確率の演算では、全ての混合多次元ガウス分布の計算において、特徴成分を同一スケールで線形量子化しているから、特徴ベクトル（浮動小数あるいは固定小数に相当する整数値）をスカラ量子化する処理は、各特徴毎に 1 フレーム当たり 1 回で良い。さらに、参照すべきデー

タとそのデータが属する中間テーブルの先頭アドレスとの差(特徴オフセット若しくはテーブルオフセット)も、特徴成分毎に共通である。したがって、単一ガウス分布の計算は、中間テーブルの先頭アドレスのロード、中間テーブルの先頭アドレスと特徴オフセットの加算、中間テーブルのアクセス、数値テーブルのアクセス、で実行できる。これにより、出力確率の演算速度を増強することができる。

適応においては数値テーブルそれ自体の書換えが不要である。ポインタテーブルを用いる場合には中間テーブルの書換えも必要ない。適応による分散や平均の変化に応じてポインタテーブル上のアクセスポインタの値だけを修正すれば済む。これにより、適応処理も高速化することができる。

数値テーブルは、外部メモリに格納することが一般的であるが、中間テーブルのアクセスによって数値テーブル上のデータアドレスを一つ得た後に、すぐに数値テーブルのアクセスを行うのではなく、多次元ガウス分布毎に全てのデータアドレスを予め求めてから、数値テーブルのアクセスを行うから、数値テーブルのアクセスを開始するまでの間に、データアドレスのデータをキャッシュメモリ 117 にプリフェッチすることができる。したがって、数値テーブルのアクセスではキャッシュヒットとなり、数値テーブルのアクセスにおけるキャッシュミス回避することができる。

以上のことから、音声認識のために出力確率を演算するとき、テーブル参照のための一連のメモリアクセスにおいて、キャッシュミスの発生なしに、3回のデータロードと(アドレス計算のための)1回の加算でガウス分布の数値が得ることができる。中間テーブルのアクセス動作が増えても、出力確率の演算を著しく高速化することができる。

また、1次元ガウス分布の分散と平均に一意に対応させて中間テーブ

ル401, 402を抽出できるグローバル中間テーブル400を採用し、グローバル中間テーブル400から抽出される中間テーブル401, 402の先頭アドレスは、ポインタテーブル420上のアクセスポインタで指定し、抽出された中間テーブルに対するアクセス位置は特徴成分を  
5 線形量子化して得られる特徴オフセットによって指定する。したがって、適応により、分散や平均が変更されても、中間テーブルの書き換えは発生せず、単に、変更に係るアクセスポインタの値をポインタテーブル上で書き換えて対応でき、適応処理の高層化も実現することができる。

また、アクセスポインタの値は分散や平均と相関が有るから、適応に  
10 よって分散や平均が変わるとき、それに応じてアクセスポインタの値を変更する処理は簡単になる。

アクセスポインタテーブルを複数セット設けておき、話者適応などによってアクセスポインタテーブルを切換えて使用することにより、話者適応の高速化を図ることができる。

15 以上本発明者によってなされた発明を実施例に基づいて具体的に説明したが本発明はそれに限定されるものではなく、その要旨を逸脱しない範囲において種々変更可能である。

例えば、データ処理システムは携帯情報端末装置に限定されない。携帯電話機能は省略してもよい。パーソナルコンピュータシステムで実行  
20 させることも可能である。

データプロセッサの構成は第2図に限定されない。データプロセッサは、マイクロプロセッサやマイクロコンピュータと称されるものを総称する。データプロセッサは命令をフェッチし、フェッチした命令を解読して演算制御処理を行う回路であり、CPU（中央処理装置）を備えて  
25 いればよい。データキャッシュメモリ或いは高速RAMを内蔵していれば更に好ましい。高速内蔵RAMにはグローバル中間テーブルやポイン

タテーブルなどを常駐させる。

- また、HMM音声認識のための出力確率演算のためのプログラムが格納されるコンピュータ読み取り可能な媒体は、フロッピーディスク、磁気テープ、ハードディスク等の磁気記憶媒体、CD-ROMやMO等の
- 5 光学的記憶媒体、メモリカード等の半導体製記録媒体、或いはそれ以外のどのような媒体であってもよい。

#### 産業上の利用可能性

- 本発明は、HMMを用いた音声認識技術に広く適用することができ、
- 10 例えば、マイクロコンピュータで制御され若しくは電池駆動される携帯情報端末装置等の実現される音声認識に適用して有効な技術に関するものである。また、本発明に係る音声認識のための出力確率の演算処理プログラムは、これをコンピュータ読み取り可能な記録媒体や通信回線などを介して、パーソナルコンピュータなどのコンピュータにロードし
- 15 て利用することも可能である。



## 請 求 の 範 囲

1. 特徴ベクトルに対してHMM音声認識を行なうためにデータプロセ  
ッサが中間テーブル及び数値テーブルを参照して混合多次元ガウス  
5 分布によって表現される出力確率を演算可能なデータ処理システム  
であって、

前記数値テーブルは、複数種類の1次元ガウス分布を基にした夫々の分布の数値を格納する領域を有し、

- 10 前記中間テーブルは、前記特徴ベクトルの特徴成分の値に対する線形量子化値に基づいて選択される領域にその量子化値に対応する前記数値テーブルの値の所在を示すためのアドレス情報を格納する領域を有し、

- 前記データプロセッサは、前記特徴成分の値を線形量子化すると共に、特徴成分毎のアクセスポイントにより中間テーブルを選択し、前  
15 記線形量子化された値を基に前記選択した中間テーブルよりアドレス情報を取得し、取得したアドレス情報を用いて数値テーブルを参照する処理を行い、数値テーブルから参照した値に基づいて前記出力確率を演算可能であることを特徴とするデータ処理システム。

2. 混合多次元ガウス分布の多次元ガウス分布毎に特徴成分毎の前記ア  
20 クセスポイントが配置されるアクセスポイントテーブルの形成領域を有し、データプロセッサは、前記アクセスポイントテーブルのアクセスポイントを用いて中間テーブルの選択を行なうものであることを特徴とする請求の範囲第1項に記載のデータ処理システム。

3. 前記1次元ガウス分布を基にした夫々の前記分布全体は2のN乗個  
25 の数値によって表現され、前記特徴成分の量子化値はその値の上位Nビットであることを特徴とする請求の範囲第1項又は第2項に記載

のデータ処理システム。

4. 前記データプロセッサは、前記数値テーブルを参照する処理を特徴成分毎に繰り返して多次元ガウス分布の値を演算し、この多次元ガウス分布の値を演算する処理を所定回数繰り返して混合多次元ガウス分布によって表現される出力確率を演算するものであることを特徴とする請求の範囲第1項又は第2項に記載のデータ処理システム。
5. 前記中間テーブルは、前記数値テーブルの基準になる1次元ガウス分布の平均位置を起点に分散の複数倍の範囲に対して前記アドレス情報を格納する領域を有し、その外側には、前記平均からの距離情報を格納する領域を有し、前記データプロセッサは、多次元ガウス分布の値を演算するために前記数値テーブルを参照する処理を特徴成分毎に繰り返していくとき、中間テーブルから参照した情報が前記距離情報であるときこれを累積し、その累積値が所定値を越えたとき、当該多次元ガウス分布のための演算を中止するものであることを特徴とする請求の範囲第4項に記載のデータ処理システム。
6. 前記中間テーブルは、前記距離情報の外側に固定値を格納する領域を有し、前記データプロセッサは、前記中間テーブルから前記固定値を参照したとき、現在処理中の当該多次元ガウス分布のための演算を中止するものであることを特徴とする請求の範囲第5項に記載のデータ処理システム。
7. 特徴ベクトルに対してHMM音声認識を行なうためにデータプロセッサがグローバルテーブル及び数値テーブルを参照して混合多次元ガウス分布によって表現される出力確率を演算可能なデータ処理システムであって、
- 前記数値テーブルは、相互に平均が同一であって分散の異なる複数種類の1次元ガウス分布を基にした夫々の分布の数値を格納する領

域を有し、

前記グローバルテーブルは、前記数値テーブルにおける分布毎のX方向の配列をY方向に複数組格納する領域有し、

5 前記X方向の配列は、前記特徴ベクトルの特徴成分の値に対する線形量子化値に基づいて選択される位置にその量子化値に対応する前記数値テーブルの値の所在を示すためのアドレス情報を格納する領域を保有し、

10 前記データプロセッサは、前記特徴成分の値を線形量子化すると共に、複数組のX方向の配列に対するY方向の選択に分散が考慮されX方向の配列に対する先頭位置の決定に平均が考慮された特徴成分毎のアクセスポインタの値に従って前記グローバルテーブルから中間テーブルを抽出し、前記抽出した中間テーブルの先頭位置を起点に、前記線形量子化値に基づいて前記アドレス情報を取得し、取得したアドレス情報を用いて数値テーブルを参照する処理を行い、数値テーブル  
15 から参照した値に基づいて前記出力確率を演算可能であることを特徴とするデータ処理システム。

20 8. 混合多次元ガウス分布の多次元ガウス分布毎に特徴成分毎の前記アクセスポインタが配置されたアクセスポインタテーブルを形成する領域有し、データプロセッサは、前記アクセスポインタテーブルのアクセスポインタを用いて中間テーブルの抽出を行なうものであることを特徴とする請求の範囲第7項に記載のデータ処理システム。

25 9. 前記データプロセッサは、適応によって混合多次元ガウス分布の平均と分散の双方又は一方が変更されるとき、これに応じて前記アクセスポインタテーブルのアクセスポインタ値を変更するものであることを特徴とする請求の範囲第8項に記載のデータ処理システム。

10. 前記アクセスポインタテーブルを複数セット形成可能な領域を有

し、前記データプロセッサは、話者を識別し、その識別結果に応じたアクセスポインタテーブルを用いるものであることを特徴とする請求の範囲第8項に記載のデータ処理システム。

5 11. 前記話者の識別は話者を明示するスイッチの状態に基づいて行なうものであることを特徴とする請求の範囲第10項に記載のデータ処理システム。

10 12. 前記アクセスポインタテーブルと話者とを対応付ける管理テーブルの形成領域を有し、前記データプロセッサは、話者の特徴を示す予め登録した識別用特徴情報と実際の音声特徴分析結果との比較結果に基づいて前記話者の識別を行ない、この識別された話者が、前記管理テーブルに登録されている話者であるときは、当該登録話者のアクセスポインタテーブルを参照するものであることを特徴とする請求の範囲第10項に記載のデータ処理システム。

15 13. 前記データプロセッサは、前記管理テーブルに登録可能な話者の人数を一定に限定すると共に、登録話者毎の使用頻度の情報を前記管理テーブルに追加し、音声特徴分析結果が登録話者であるとき、分析結果に一致する登録話者の使用頻度をインクリメントし、分析結果に不一致の登録話者の使用頻度をデクリメントし、音声特徴分析結果が登録話者以外であるときは、最低使用頻度の登録話者を前記管理テーブルから削除し、これに代えて当該登録話者以外の話者を管理テーブルに追加するものであることを特徴とする請求項12に記載のデータ処理システム。

20 14. 複数の音声入力系を持ち、個々の音声入力系毎に、前記アクセスポインタテーブルの形成領域を有し、前記データプロセッサは、前記複数の音声入力系に対して、独立にアクセスポインタテーブルを用いて、並列的な音声認識を可能にするものであることを特徴とする請求

の範囲第 8 項に記載のデータ処理システム。

- 1 5 . 前記データプロセッサは、特徴ベクトルの全ての特徴成分に対して、線形量子化を行なうと共に、前記量子化値と前記 X 方向の配列の単一配列要素のアドレス量との積に基づいて、抽出される中間テーブルの先頭位置からの特長オフセットを算出し、その後、多次元混合ガウス分布毎に、前記アクセスポイントと特徴オフセットとにより中間テーブルの参照を行って数値テーブルの参照を行うものであることを特徴とする請求の範囲第 7 項又は第 8 項に記載のデータ処理システム。
- 10 1 6 . 前記 1 次元ガウス分布を基にした夫々の前記分布全体は 2 の N 乗個の数値によって表現され、前記特徴成分の量子化値はその値の上位 N ビットであることを特徴とする請求の範囲第 1 5 項に記載のデータ処理システム。
- 15 1 7 . 前記データプロセッサは、前記数値テーブルを参照する処理を特徴成分毎に繰り返して多次元ガウス分布の値を演算し、この多次元ガウス分布の値を演算する処理を所定回数繰り返して混合多次元ガウス分布によって表現される出力確率を演算するものであることを特徴とする請求の範囲第 1 6 項に記載のデータ処理システム。
- 20 1 8 . 前記 X 方向の夫々の配列は、前記数値テーブルの基準になる 1 次元ガウス分布の平均位置を起点に分散の複数倍の範囲に対して前記アドレス情報を格納する領域を有し、その外側には、前記平均からの距離情報を格納する領域を有し、前記データプロセッサは、多次元ガウス分布の値を演算するために前記数値テーブルを参照する処理を特徴成分毎に繰り返していくとき、中間テーブルから参照した情報が前記距離情報であるときこれを累積し、その累積値が所定値を越えたとき、当該多次元ガウス分布のための演算を中止するものであること
- 25

を特徴とする請求の範囲第 17 項に記載のデータ処理システム。

19. 前記 Y 方向の夫々の配列は、前記距離情報の外側に固定値を格納する領域有し、前記データプロセッサは、前記中間テーブルから前記固定値を参照したとき、現在処理中の当該多次元ガウス分布のための演算を中止するものであることを特徴とする請求の範囲第 18 項に記載のデータ処理システム。

20. 特徴ベクトルに対して HMM 音声認識を行なうために、

複数種類の 1 次元ガウス分布を基にした夫々の分布の数値を保有する数値テーブルと、

- 10 前記特徴ベクトルの特徴成分の値に対する線形量子化値に基づいて選択される領域にその量子化値に対応する前記数値テーブルの値の所在を示すためのアドレス情報を保有する前記中間テーブルと、を用い、

- 15 前記特徴成分の値を線形量子化すると共に、特徴成分毎のアクセスポイントにより中間テーブルを選択し、前記線形量子化された値を基に前記選択した中間テーブルよりアドレス情報を取得し、取得したアドレス情報を用いて数値テーブルを参照する処理を行い、数値テーブルから参照した値に基づいて、混合多次元ガウス分布によって表現される出力確率を演算することを特徴とする混合ガウス HMM の出力確率演算方法。

21. 前記中間テーブルの選択は、混合多次元ガウス分布の多次元ガウス分布毎に特徴成分毎の前記アクセスポイントが配置されたアクセスポイントテーブルを用いて行なうものであることを特徴とする請求の範囲第 20 項に記載の混合ガウス HMM の出力確率演算方法。

- 25 22. 特徴ベクトルに対して HMM 音声認識を行なうために、

相互に平均が同一であって分散の異なる複数種類の 1 次元ガウス

分布を基にした夫々の分布の数値を保有する数値テーブルと、

前記数値テーブルにおける分布毎のX方向の配列をY方向に複数組保有し、前記X方向の配列には、前記特徴ベクトルの特徴成分の値に対する線形量子化値に基づいて選択される位置にその量子化値に対応する前記数値テーブルの値の所在を示すためのアドレス情報を保有するグローバルテーブルと、を用い、

前記特徴成分の値を線形量子化すると共に、複数組のX方向の配列に対するY方向の選択に分散が考慮されX方向の配列に対する先頭位置の決定に平均が考慮された特徴成分毎のアクセスポイントの値に従って前記グローバルテーブルから中間テーブルを抽出し、前記抽出した中間テーブルの先頭位置を起点に、前記線形量子化値に基づいて前記アドレス情報を取得し、取得したアドレス情報を用いて数値テーブルを参照する処理を行い、数値テーブルから参照した値に基づいて、混合多次元ガウス分布によって表現される出力確率を演算することを特徴とする混合ガウスHMMの出力確率演算方法

2 3 . 前記中間テーブルの抽出は、混合多次元ガウス分布の多次元ガウス分布毎に特徴成分毎の前記アクセスポイントが配置されたアクセスポイントテーブルを用いて行なうものであることを特徴とする請求の範囲第2 2 項に記載の混合ガウスHMMの出力確率演算方法。

2 4 . 適応によって混合多次元ガウス分布の平均と分散の双方又は一方が変更されるとき、これに応じて前記アクセスポイントテーブルのアドレスポイント値を変更するものであることを特徴とする請求の範囲第2 3 項に記載の混合ガウスHMMの出力確率演算方法。

2 5 . 特徴ベクトルを入力してHMM音声認識を行なうために、複数種類の1次元ガウス分布を基にした夫々の分布の数値を保有する数値テーブルと、

前記特徴ベクトルの特徴成分の値に対する線形量子化値に基づいて選択される領域にその量子化値に対応する前記数値テーブルの値の所在を示すためのアドレス情報を保有する前記中間テーブルと、

5 混合多次元ガウス分布の多次元ガウス分布毎に特徴成分毎のアクセスポイントが配置されたアクセスポイントテーブルと、を用い、

前記特徴成分の値を線形量子化すると共に、前記アクセスポイントテーブル上の特徴成分毎のアクセスポイントにより中間テーブルを選択し、前記線形量子化された値を基に前記選択した中間テーブルよりアドレス情報を取得し、取得したアドレス情報を用いて数値テーブルを参照する処理を行い、数値テーブルから参照した値に基づいて、  
10 混合多次元ガウス分布によって表現される出力確率を演算する処理を、コンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

26. 特徴ベクトルを入力してHMM音声認識を行なうために、

15 相互に平均が同一であって分散の異なる複数種類の1次元ガウス分布を基にした夫々の分布の数値を保有する数値テーブルと、

前記数値テーブルにおける分布毎のX方向の配列をY方向に複数組保有し、前記X方向の配列には、前記特徴ベクトルの特徴成分の値に対する線形量子化値に基づいて選択される位置にその量子化値に対応する前記数値テーブルの値の所在を示すためのアドレス情報を保有するグローバルテーブルと、  
20

複数組のX方向の配列に対するY方向の選択に分散が考慮されX方向の配列に対する先頭位置の決定に平均が考慮された特徴成分毎のアクセスポイントが混合多次元ガウス分布の多次元ガウス分布毎に配置されたアクセスポイントテーブルと、を用い、  
25

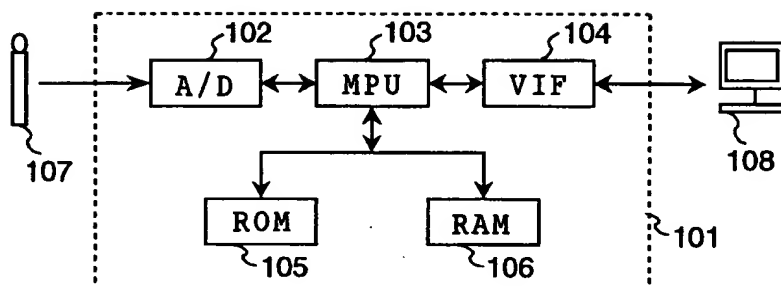
前記特徴成分の値を線形量子化すると共に、前記アクセスポイント



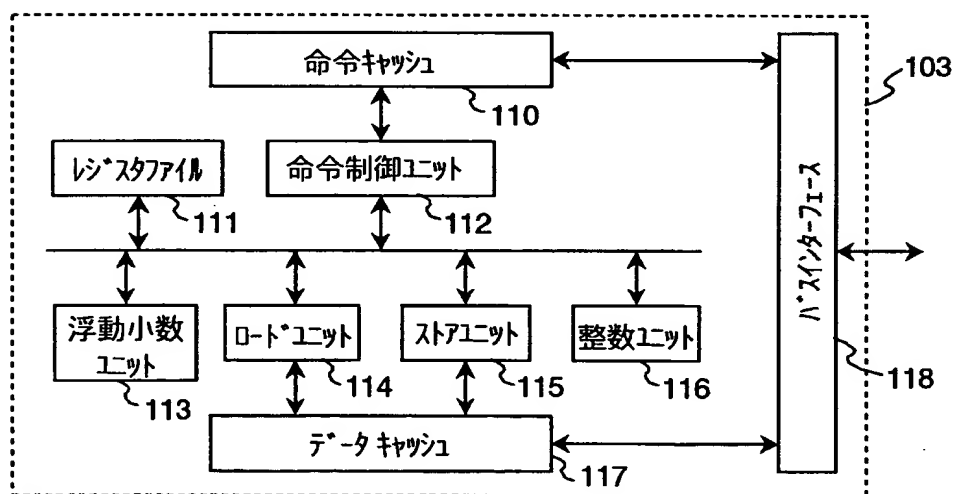
- テーブル上のアクセスポインタの値に従って前記グローバルテーブルから中間テーブルを抽出し、前記抽出した中間テーブルの先頭位置を起点に、前記線形量子化値に基づいて前記アドレス情報を取得し、取得したアドレス情報を用いて数値テーブルを参照する処理を行い、
- 5 数値テーブルから参照した値に基づいて、混合多次元ガウス分布によって表現される出力確率を演算する処理を、コンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。
27. 前記プログラムは、適応によって混合多次元ガウス分布の平均と分散の双方又は一方が変更されるとき、これに応じて前記アクセスポインタテーブル上のアクセスポインタの値を変更するものであることを特徴とする請求の範囲第23項に記載のコンピュータ読み取り可能な記録媒体。
- 10
28. 動作電源を供給するバッテリーを有し、前記データプロセッサは前記バッテリーを動作電源とし消費電力が1W以下であることを特徴とする請求の範囲第1項又は第7項に記載のデータ処理システム。
- 15

1/28

第1図

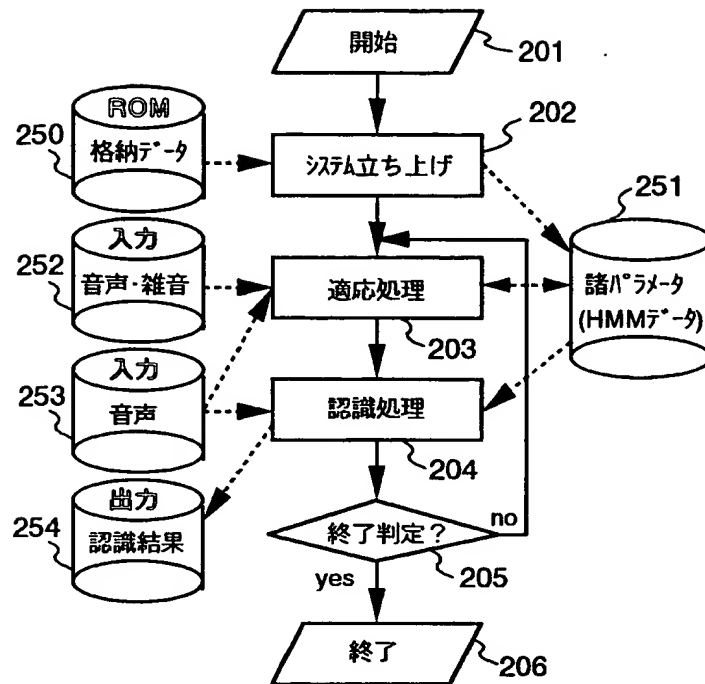


第2図

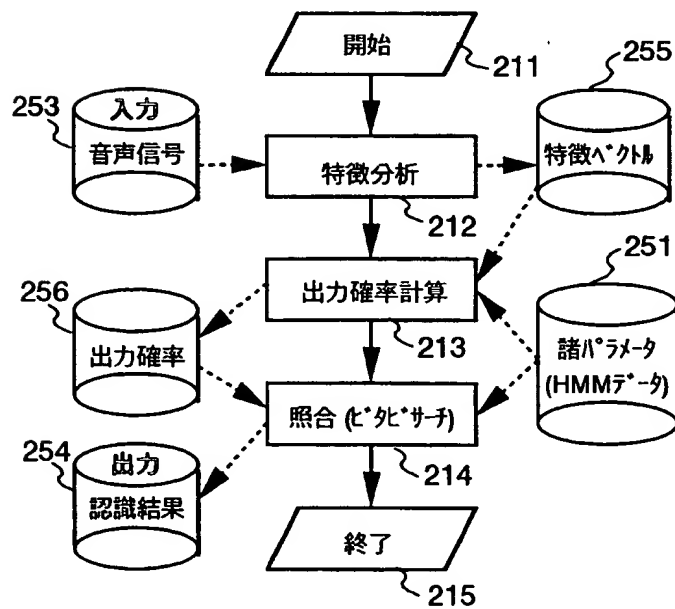


2/28

第3図

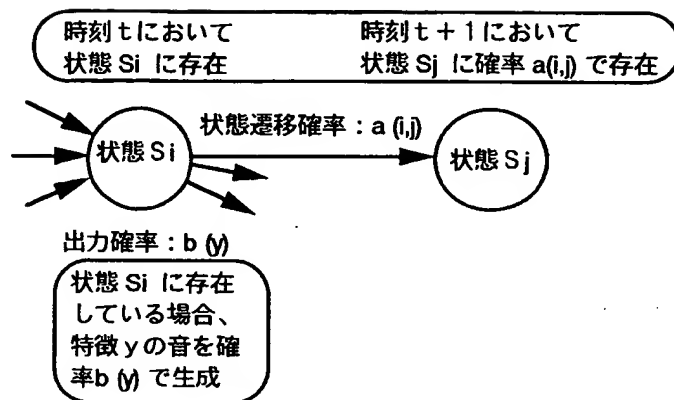


第4図

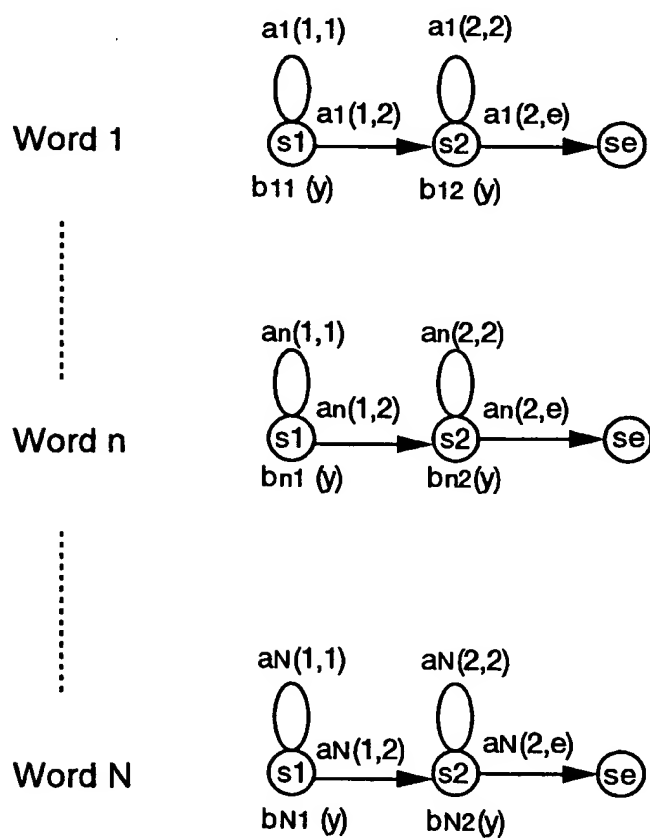


3/28

第5図

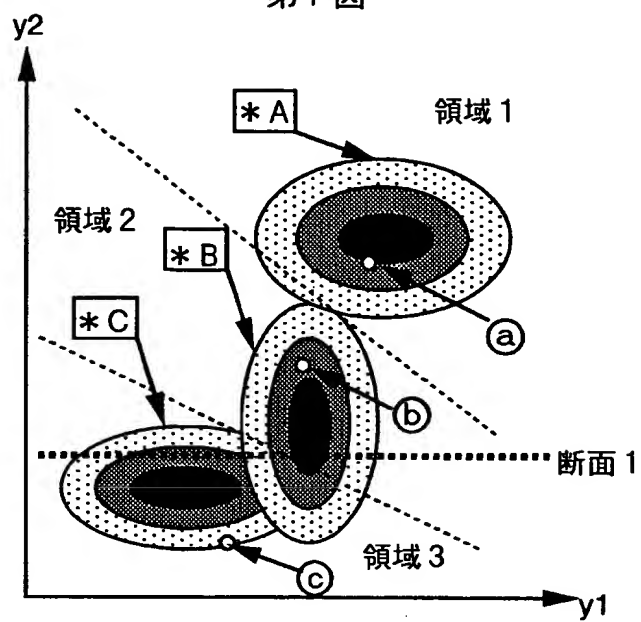


第6図

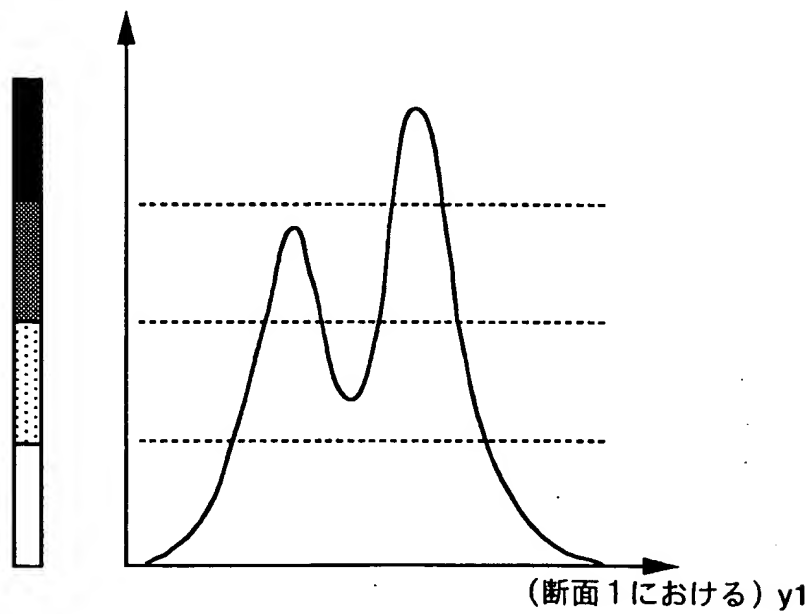


4/28

第7図

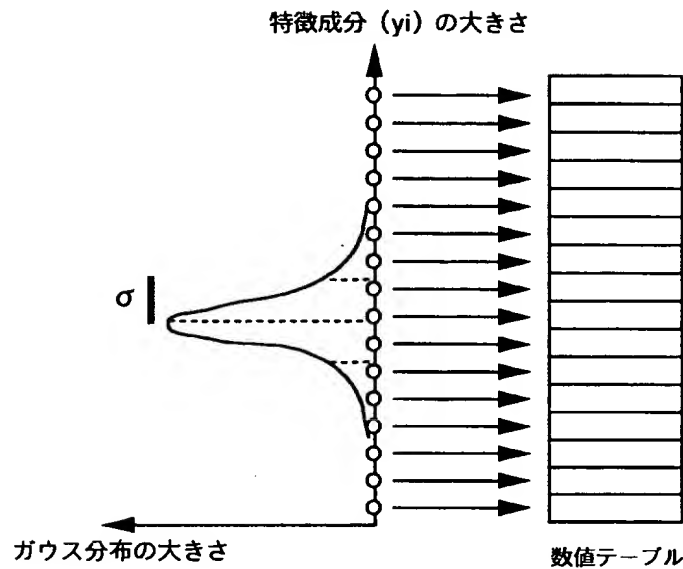


第8図

(断面 1 における)  $b(y)$ 

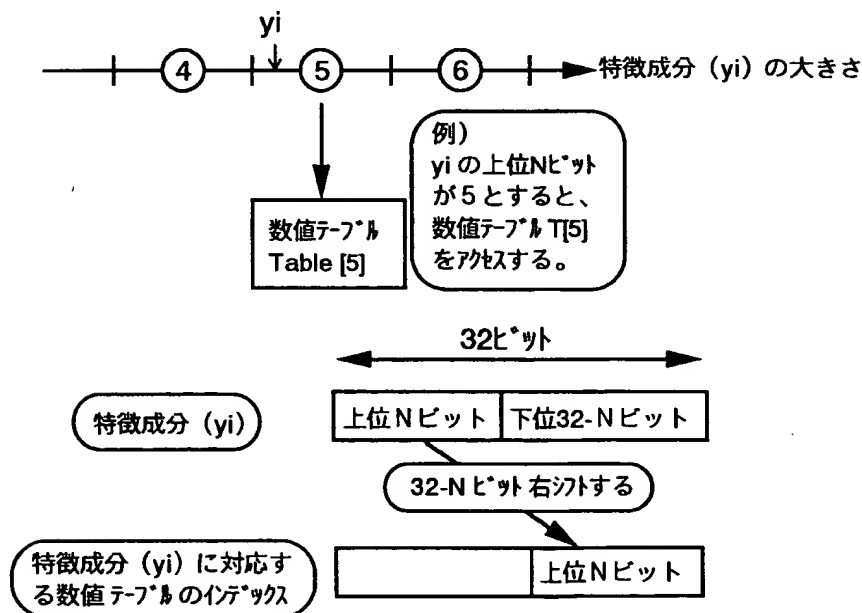
5/28

第9図



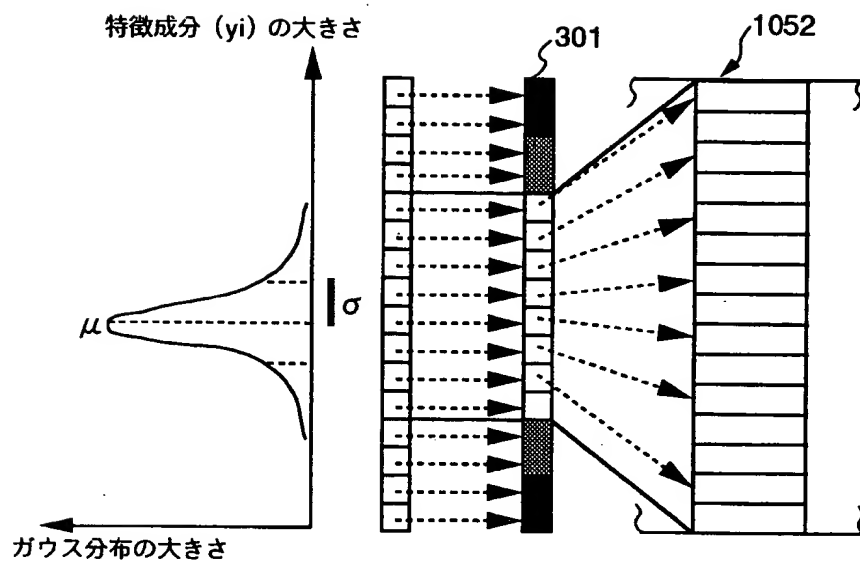
one-dimensional  
distribution + numerical  
table when  
scalar quantization.

第10図



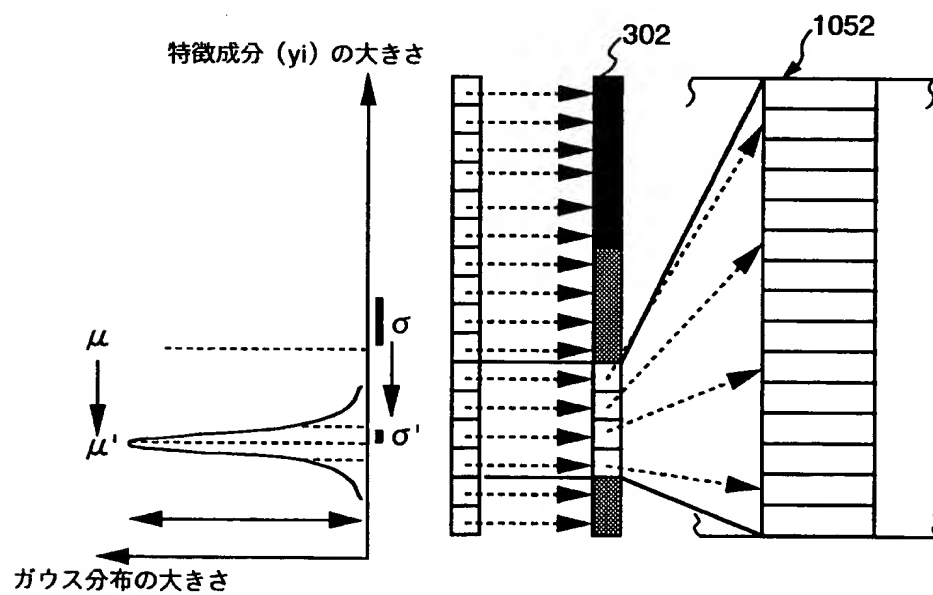
6/28

第11図



[適応前]

第12図

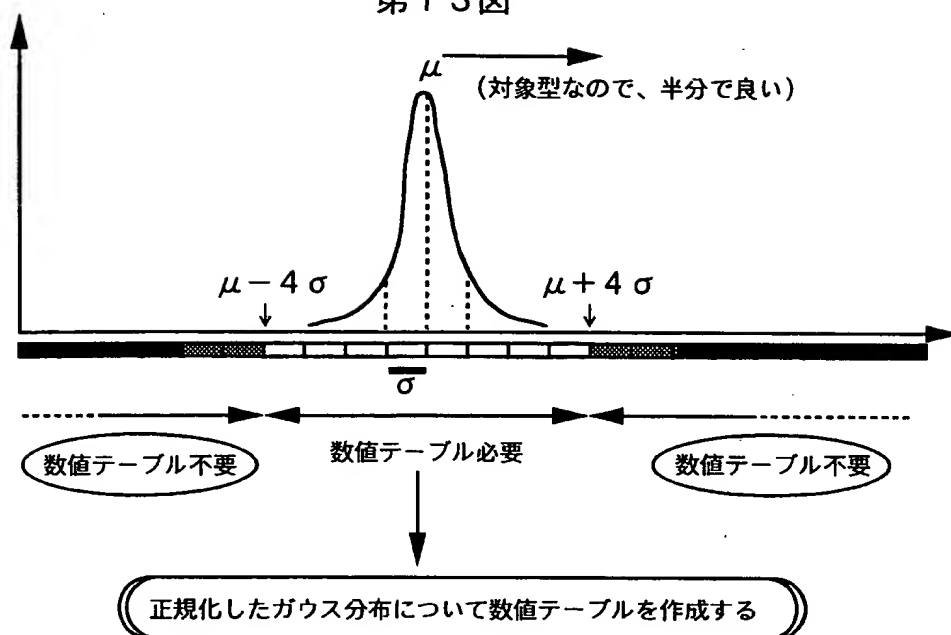


[適応後]

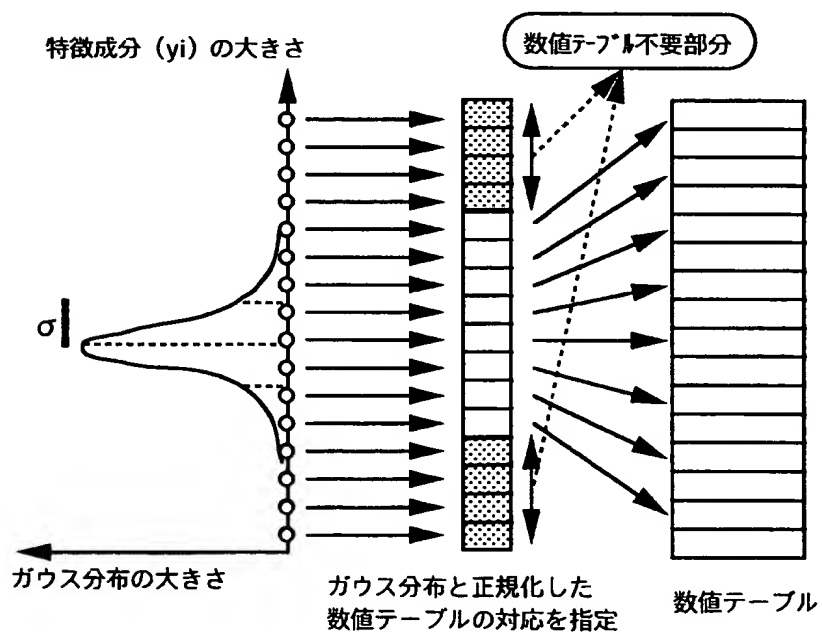
*average, variance*

7/28

第13図



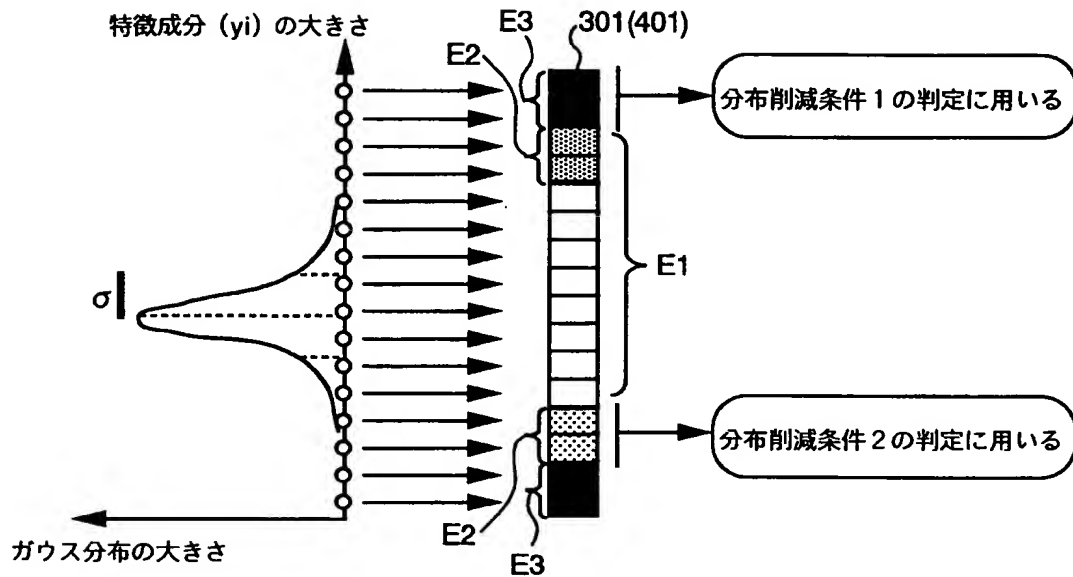
第14図



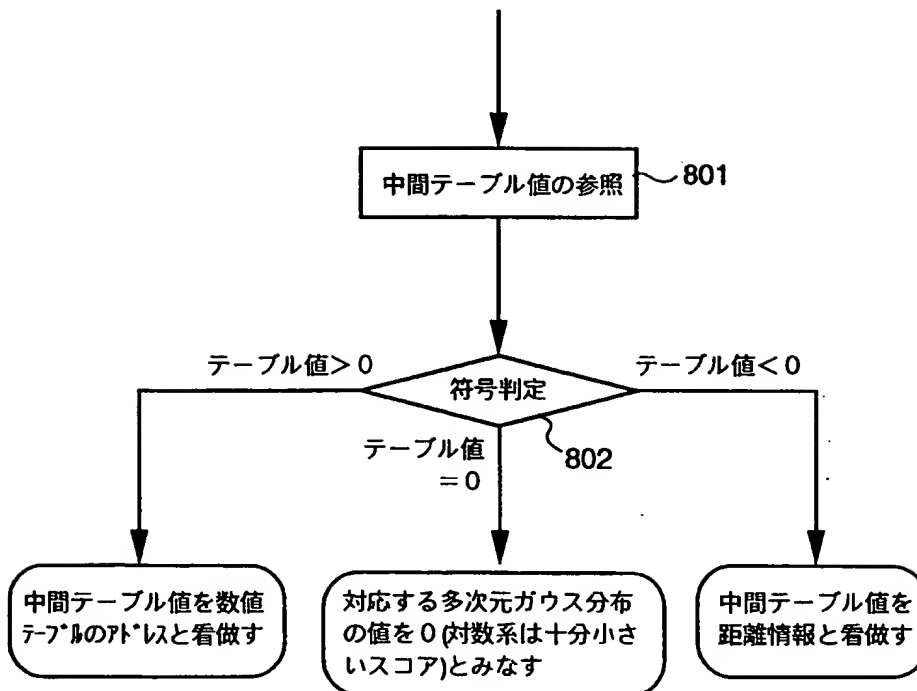


8/28

第15図



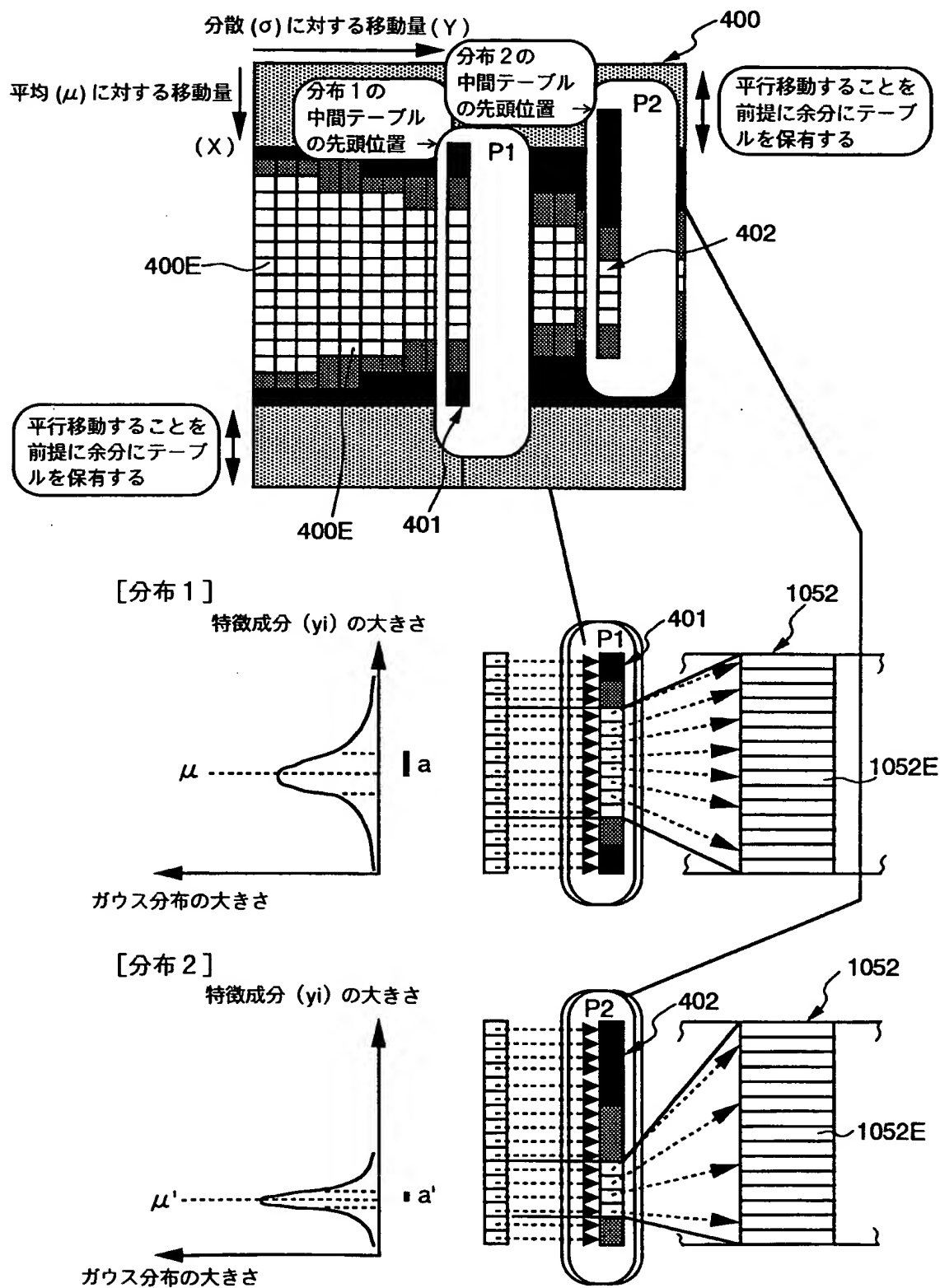
第16図



Global table

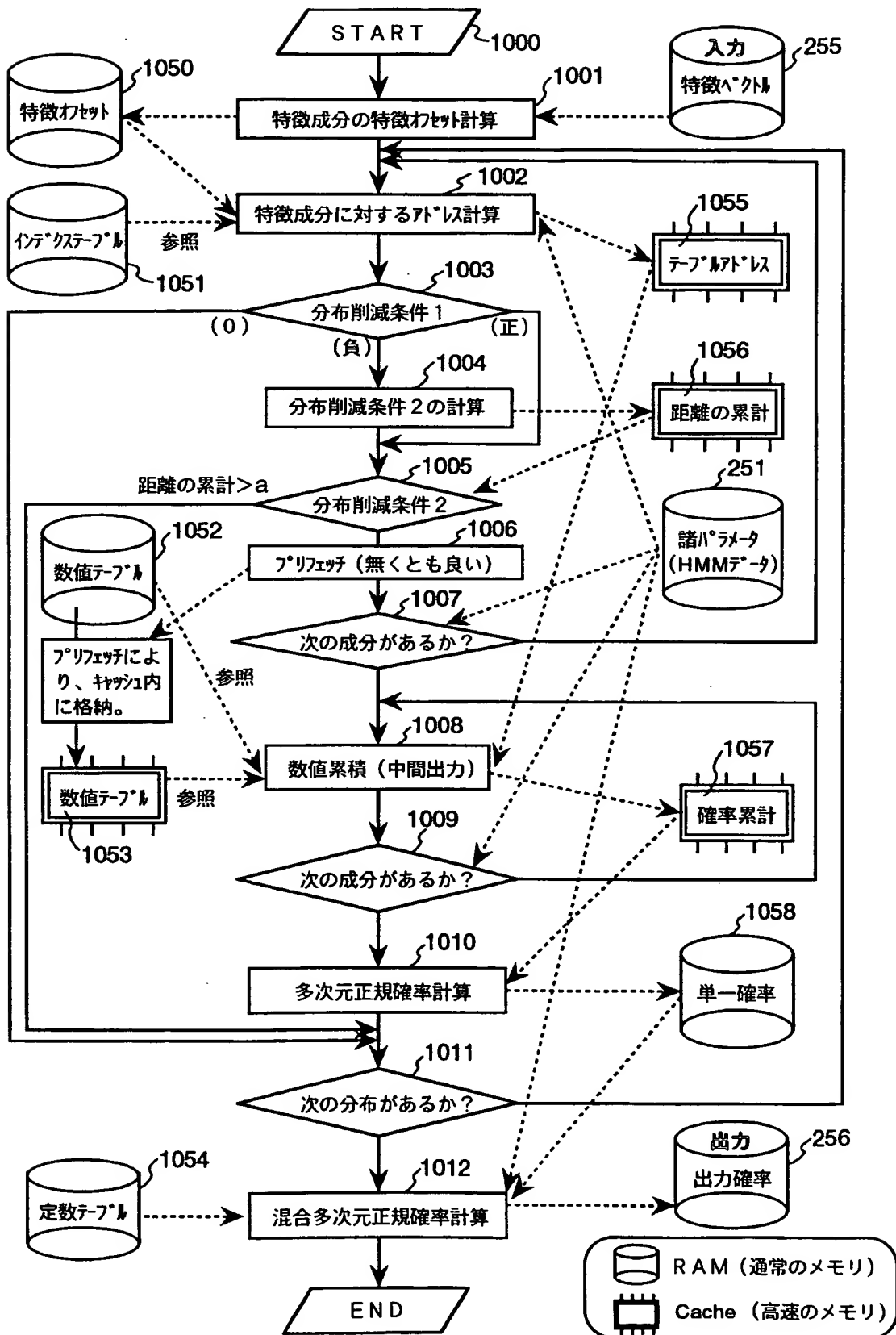
9/28

第17図



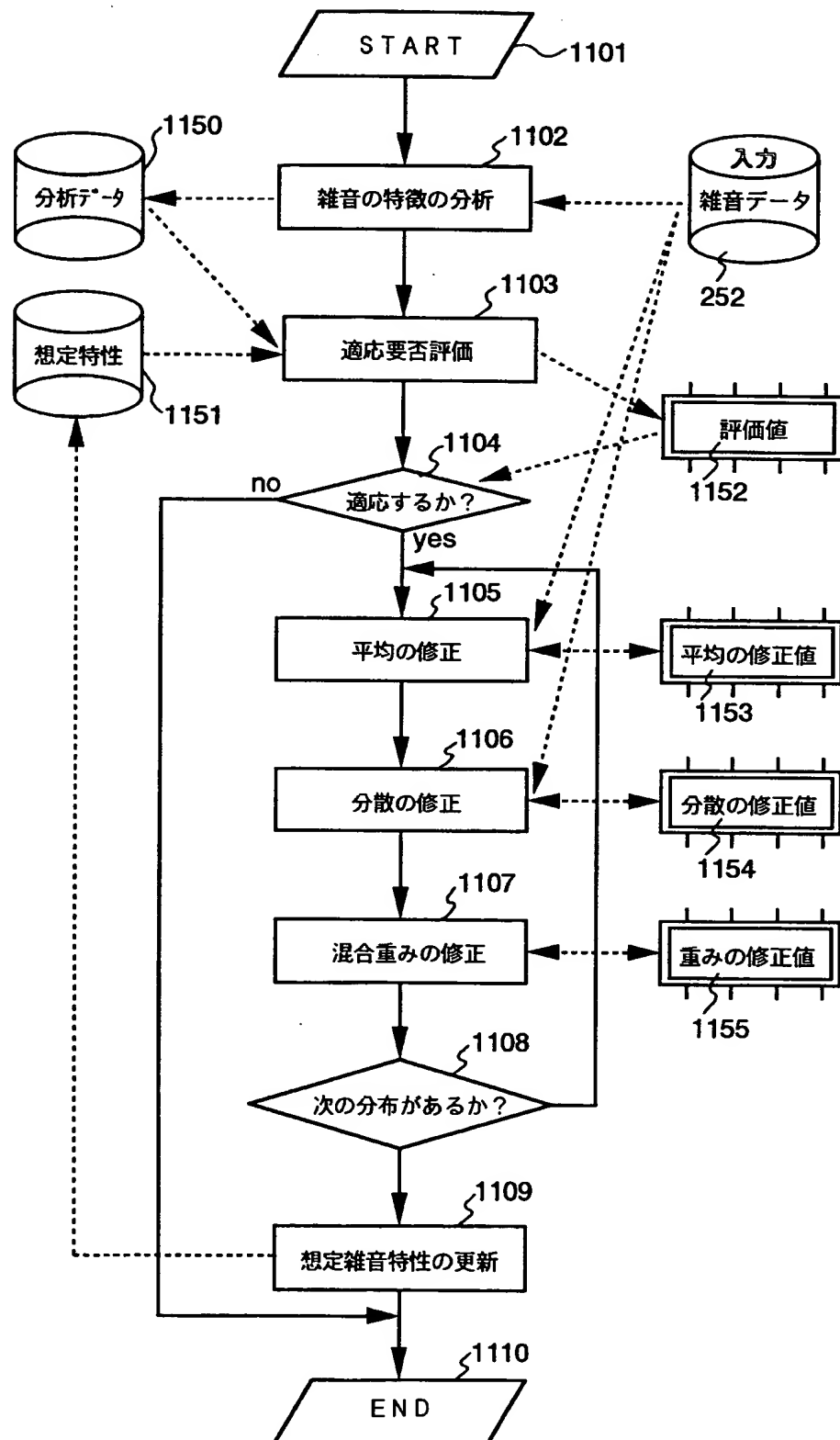
10/28

第18図



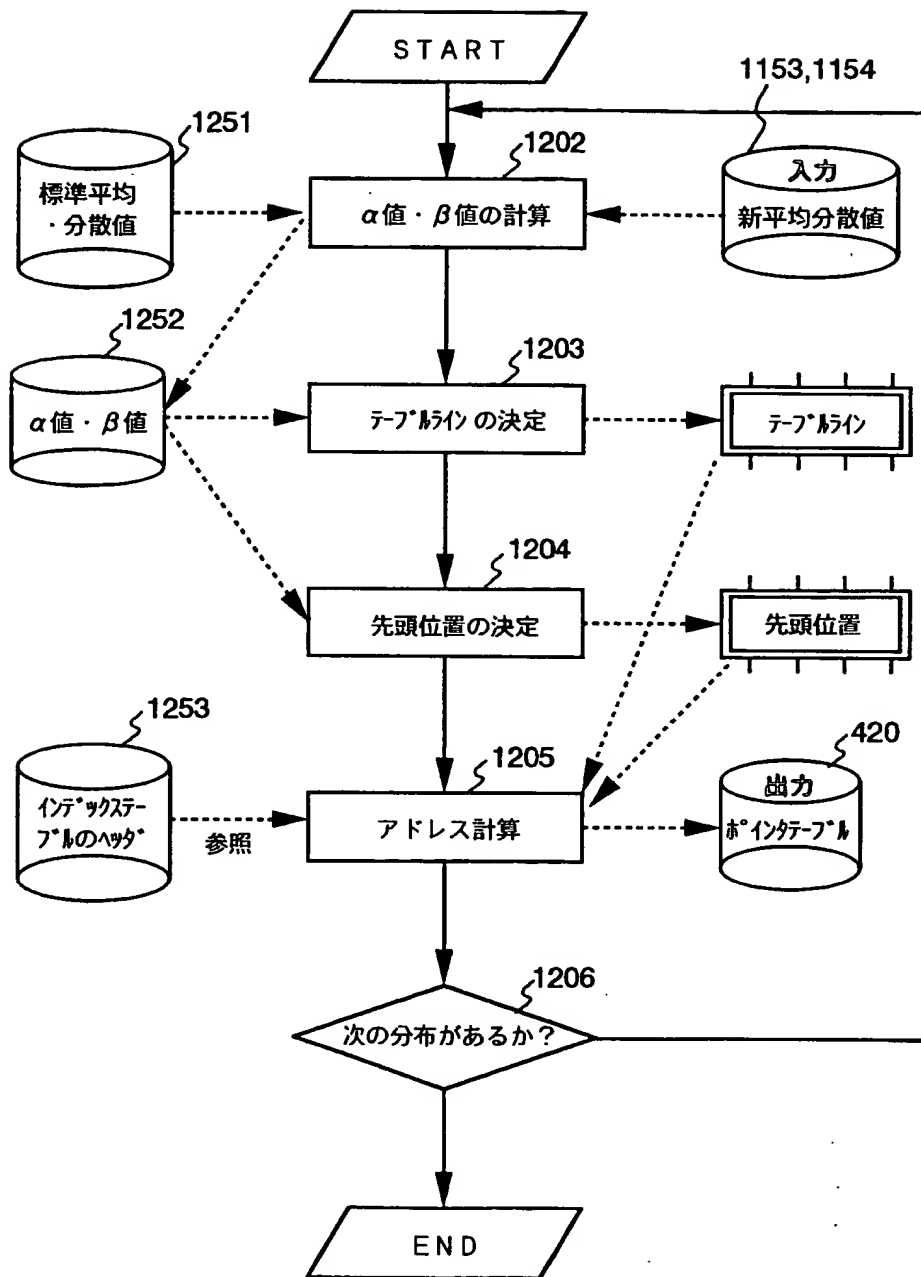
11/28

第19図



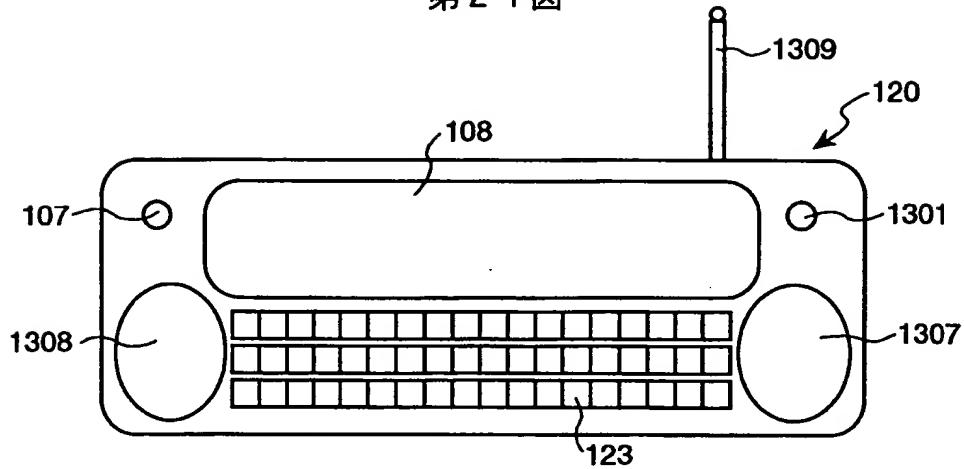
12/28

第20図

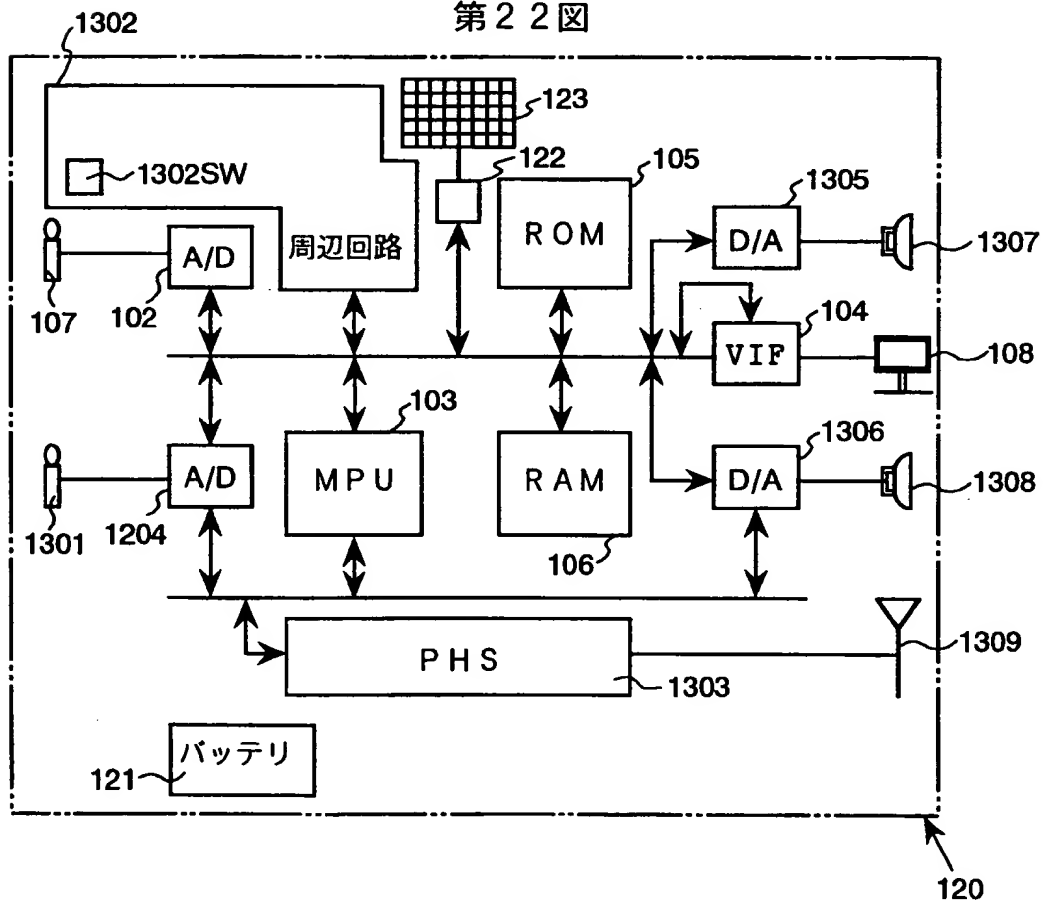


1 3 / 2 8

第 2 1 図

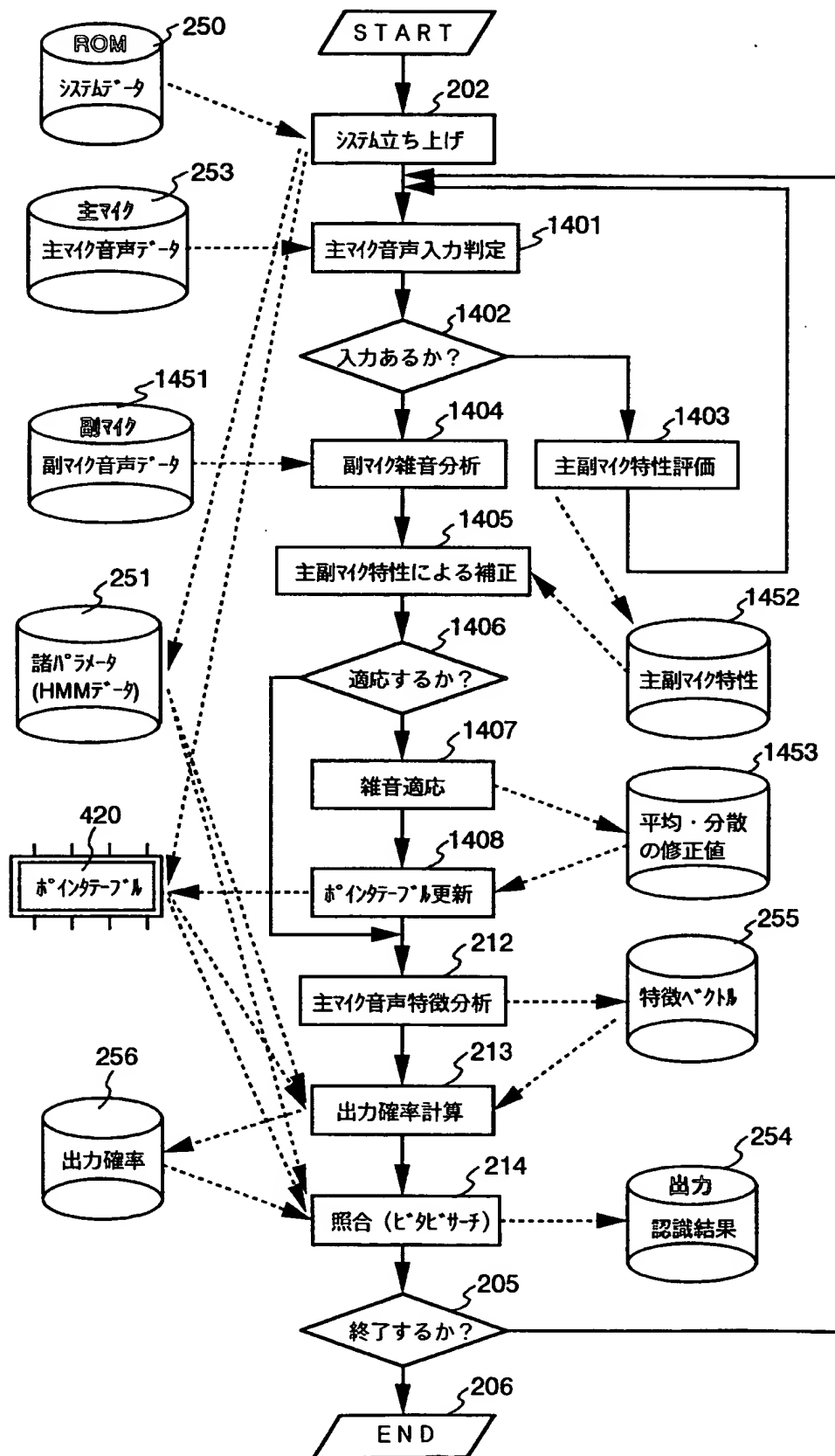


第 2 2 図



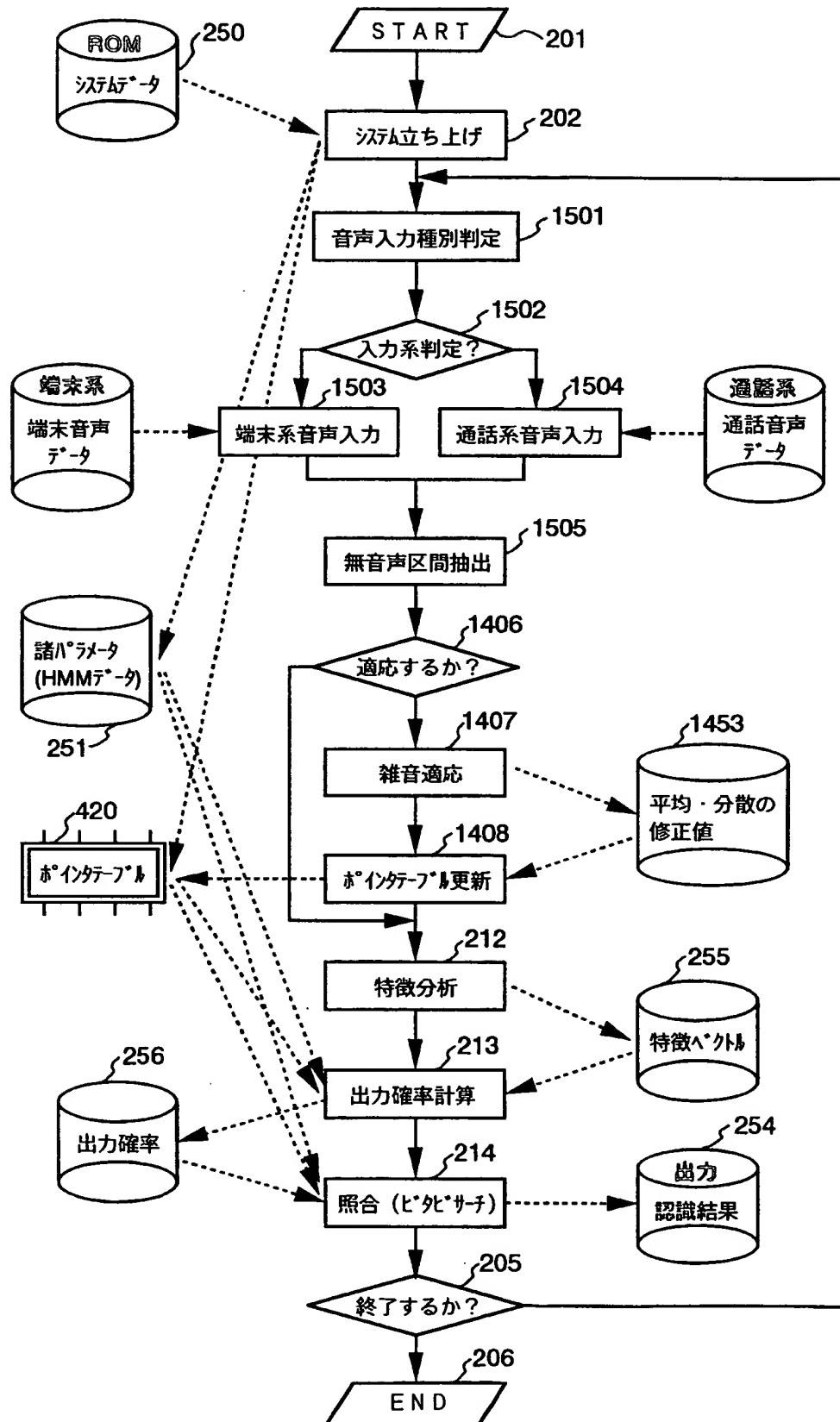
14/28

## 第23図



15/28

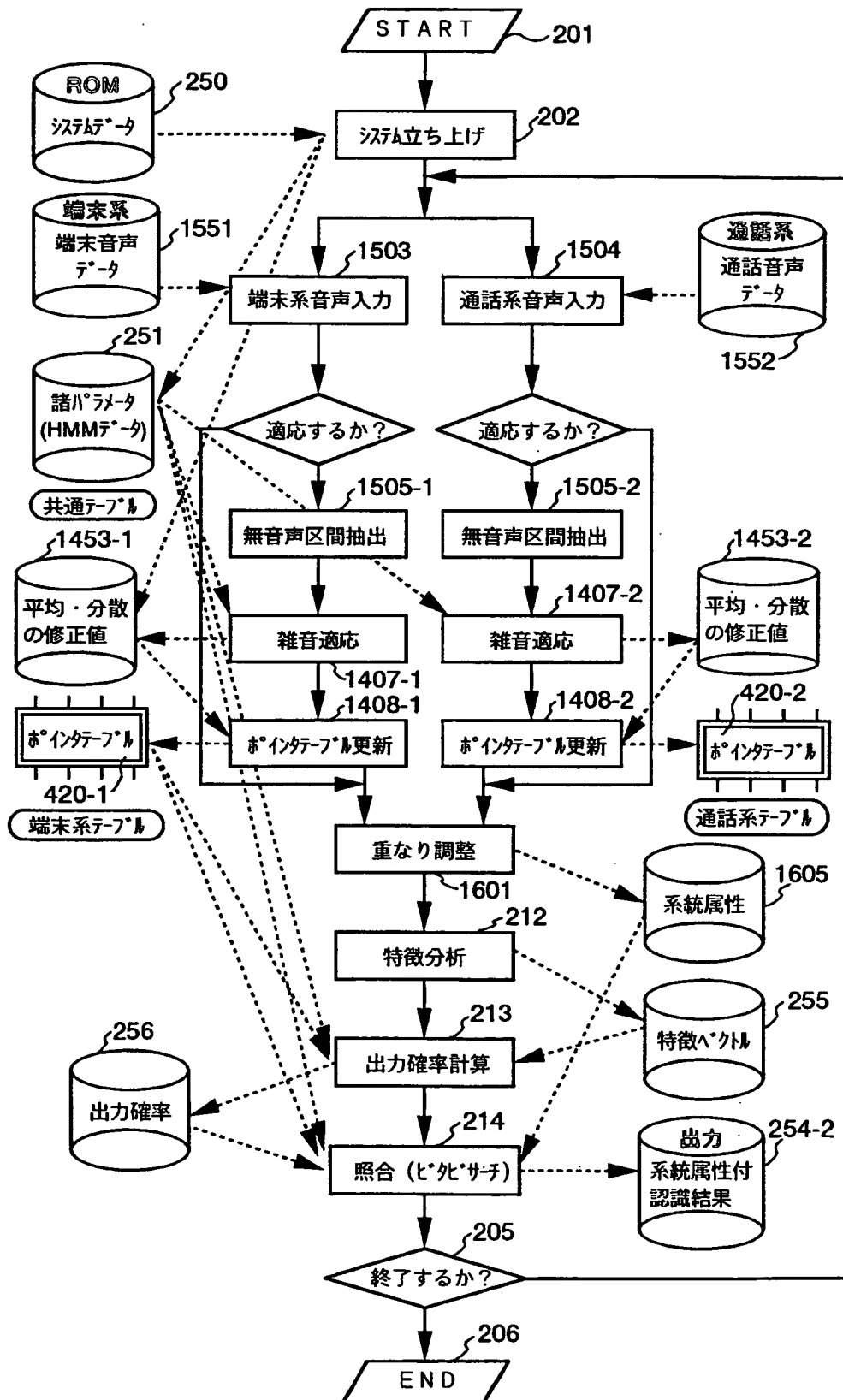
## 第24図





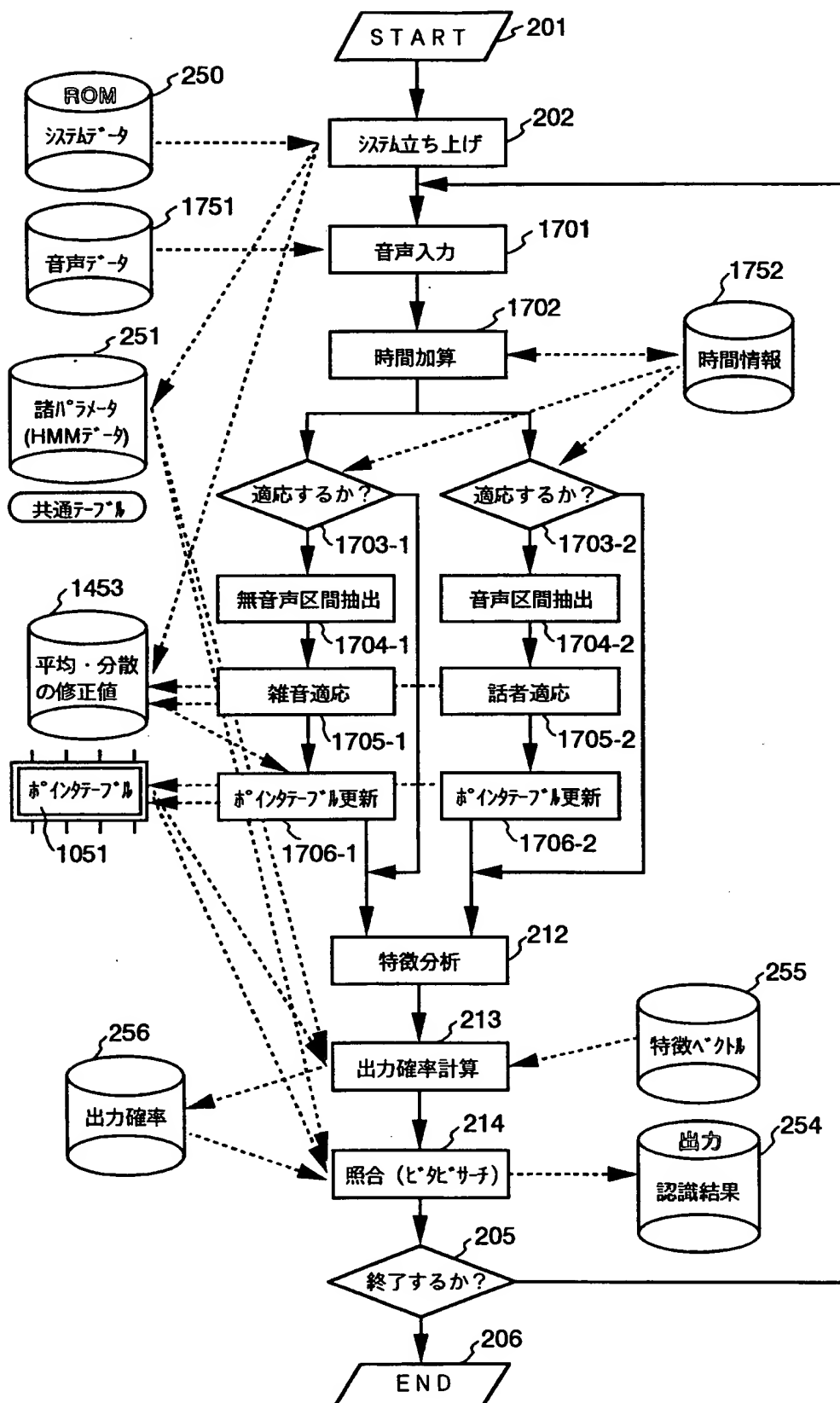
16/28

第25図



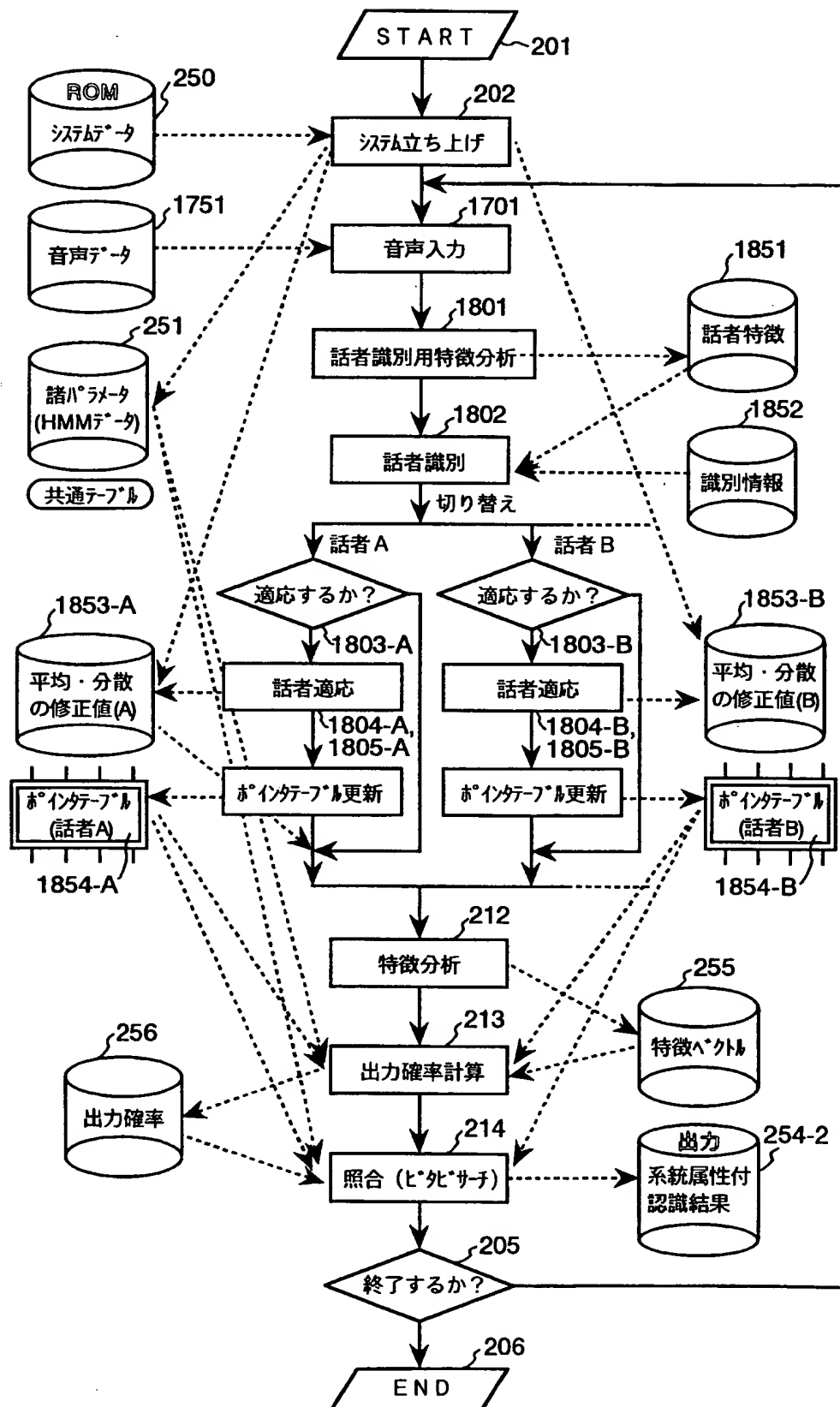
17/28

第26図



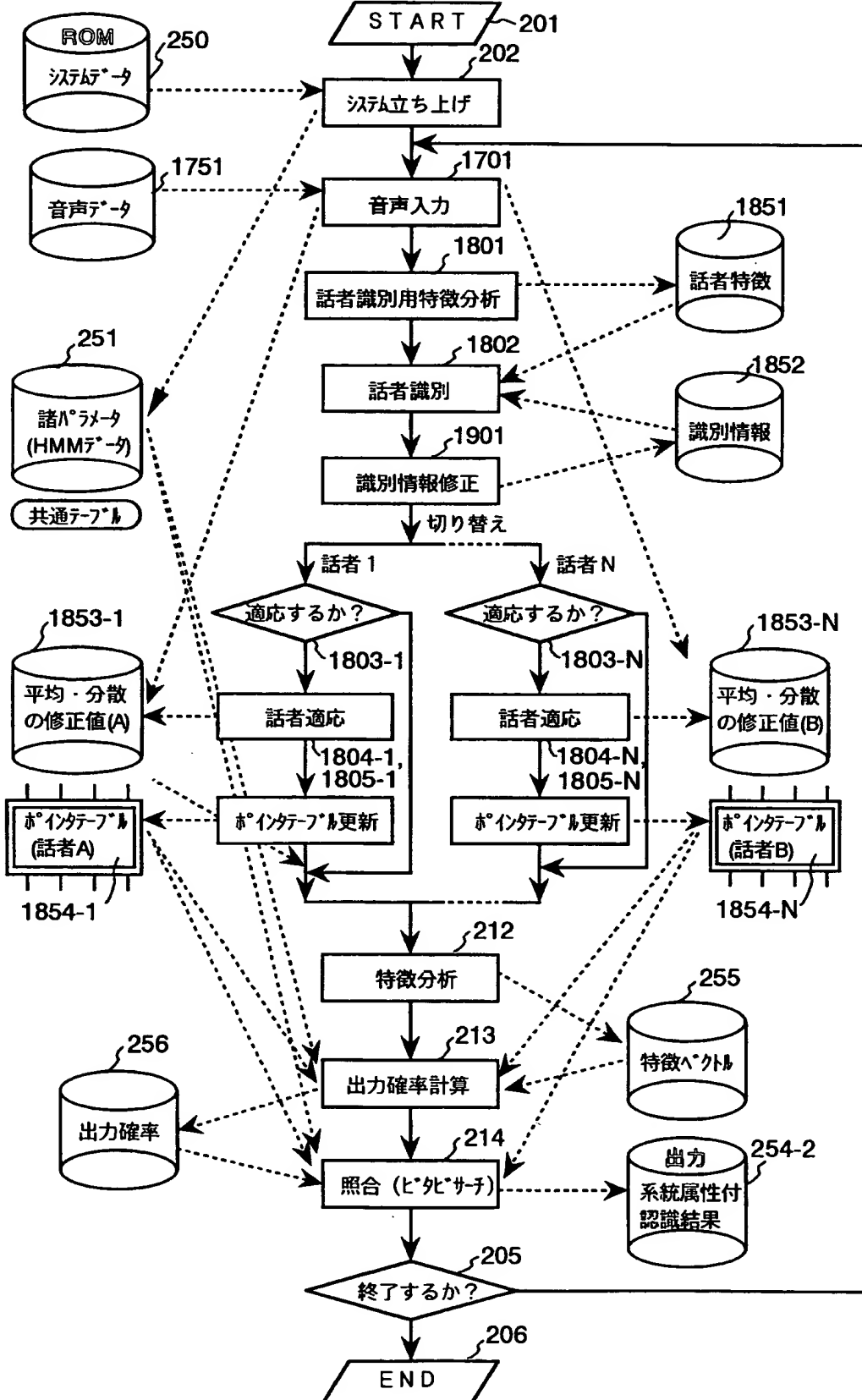
18/28

第27図



19/28

第28図

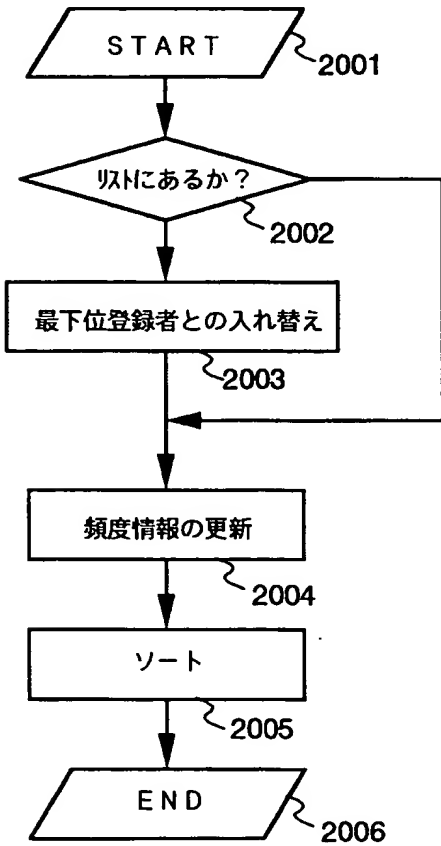


第29図

501	502	503
登録話者	頻度情報	データイン
No.5	15	-----
No.2	13	-----
-----	-----	-----
No.4	2	-----

500

第30図



21/28

第31図

登録話者	頻度情報	メモ・コメント
No.5	15	-----
No.2	13	-----
-----	-----	-----
No.4	2	-----

一番下からバブルソートする

500

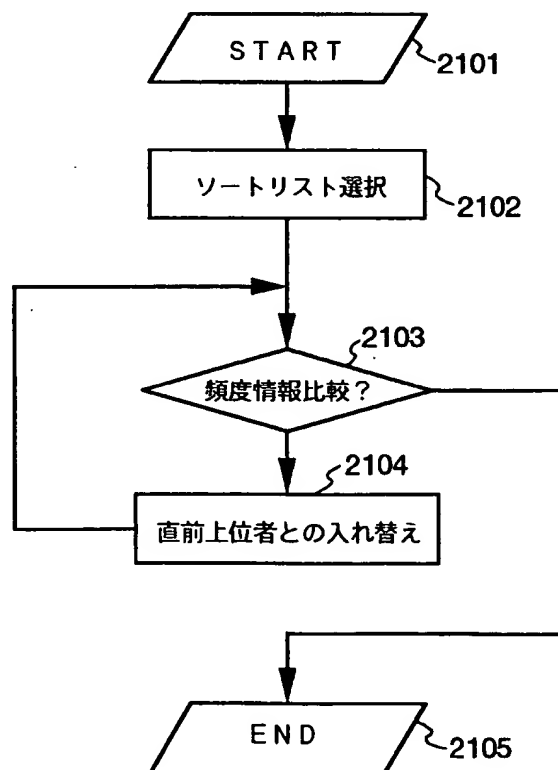
第32図

登録話者	頻度情報	メモ・コメント
No.5	15	-----
No.2	13	-----
-----	-----	-----
No.4	2	-----

リストのある所からバブルソートする

500

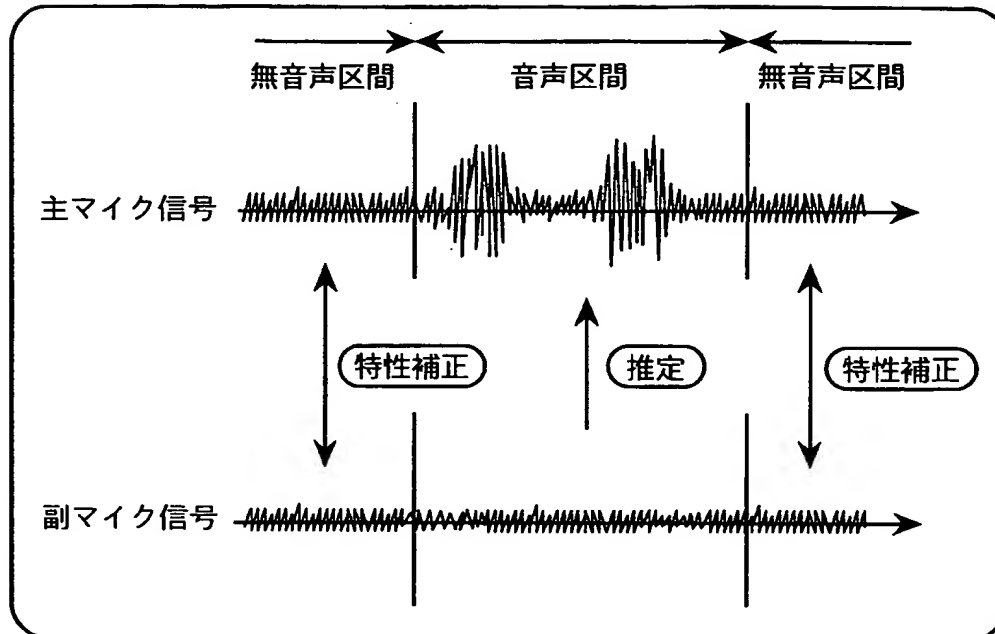
第33図



22/28

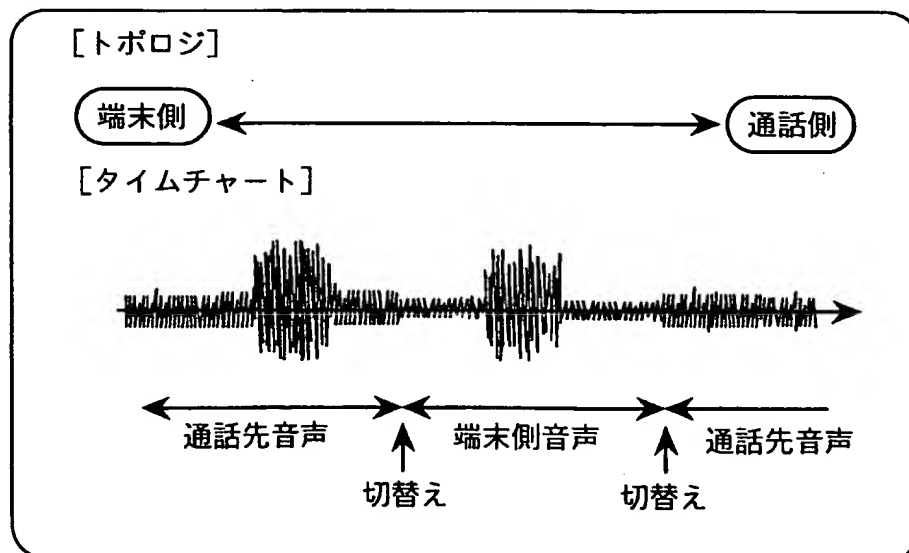
## 第34図

&lt;2マイク型雑音適応の考え方&gt;



## 第35図

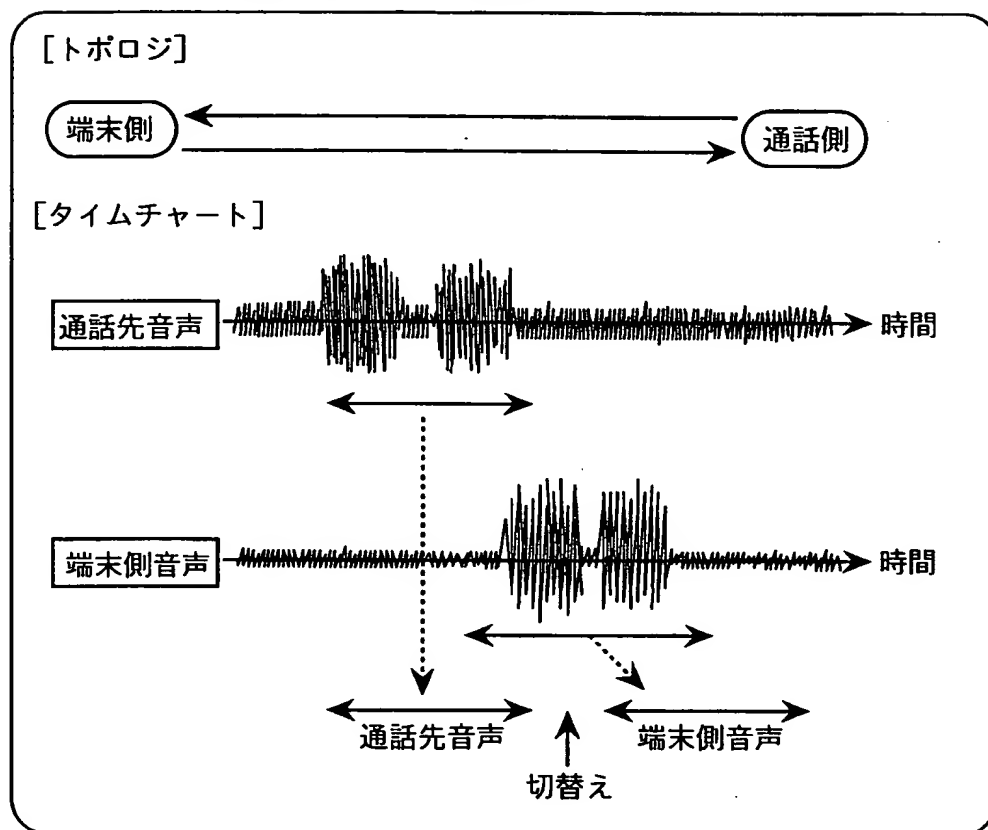
&lt;トランシーバ型通話における音声認識&gt;



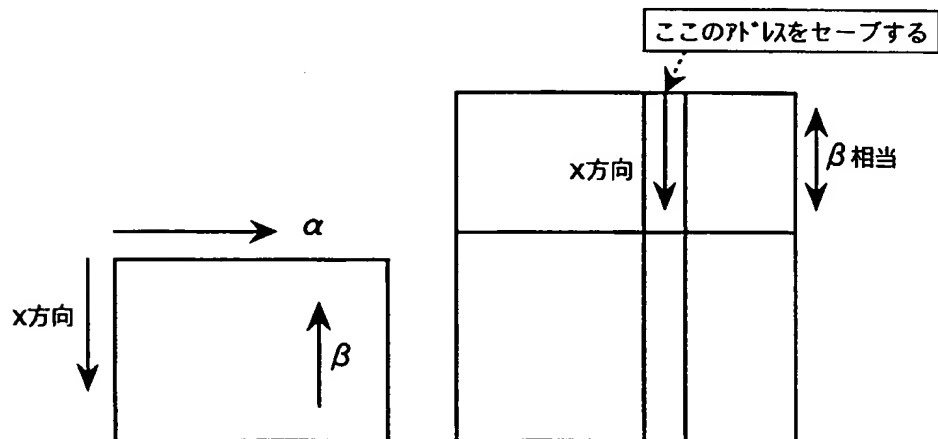
23/28

第36図

&lt;セパレート型通話における音声認識&gt;



第37図

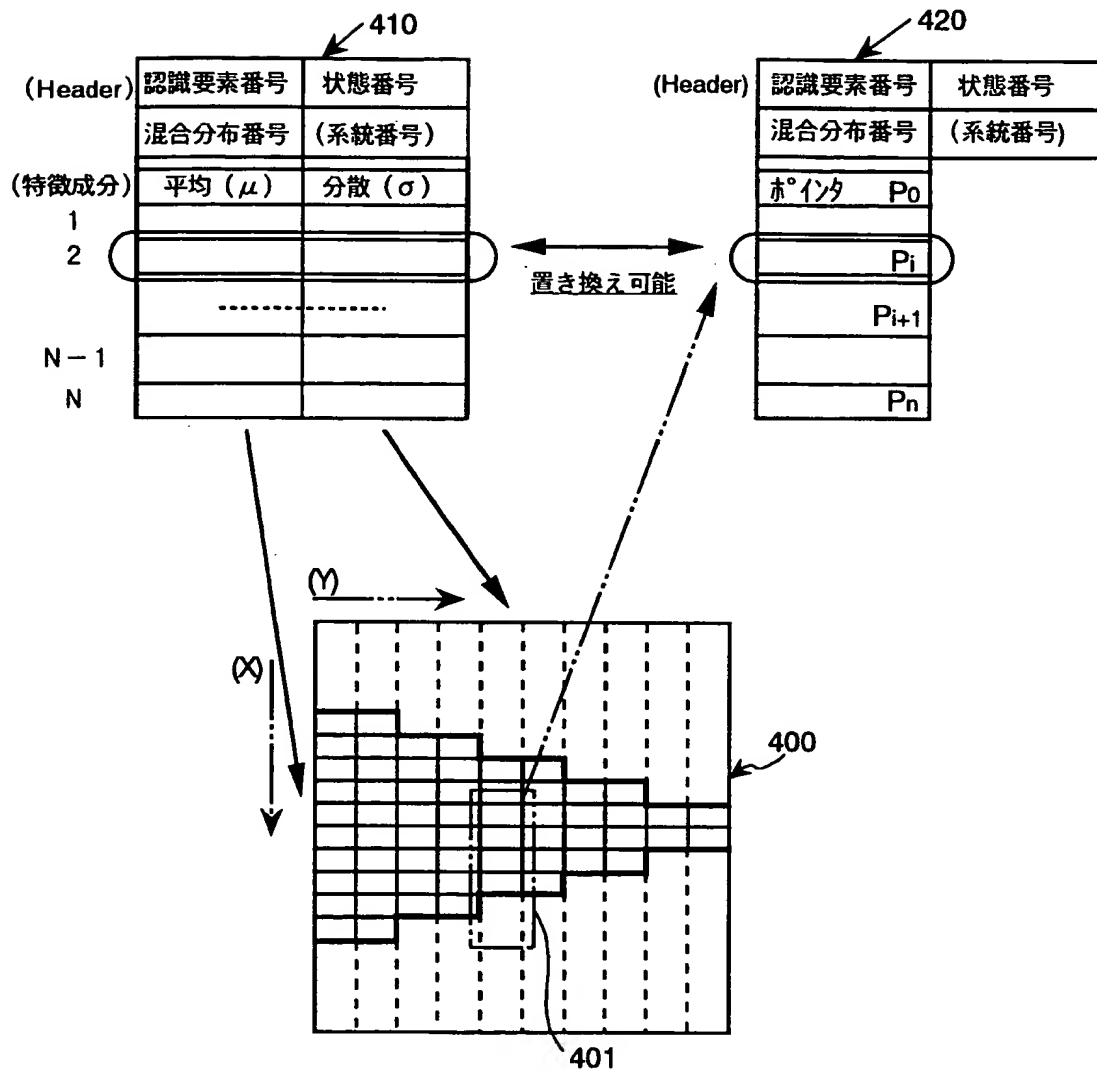




24/28

Pointer for Global Table

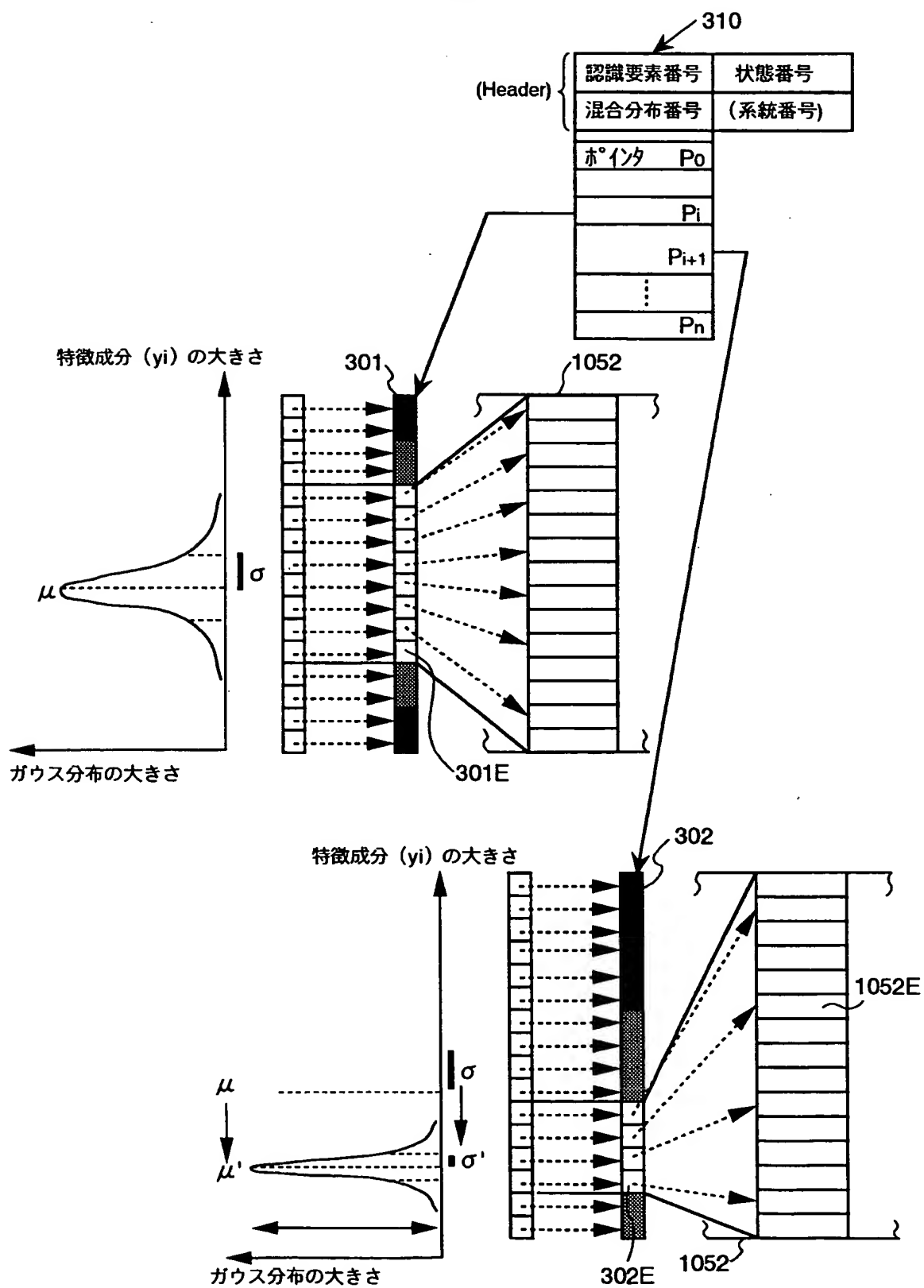
第38図



( $\mu$ ,  $\sigma$ ) のセットに対して、 $P$ は一意に決定する  
 $P$ に対して、( $\mu$ ,  $\sigma$ ) のセットは一意に決定する

25/28

第39図



26/28

第40図

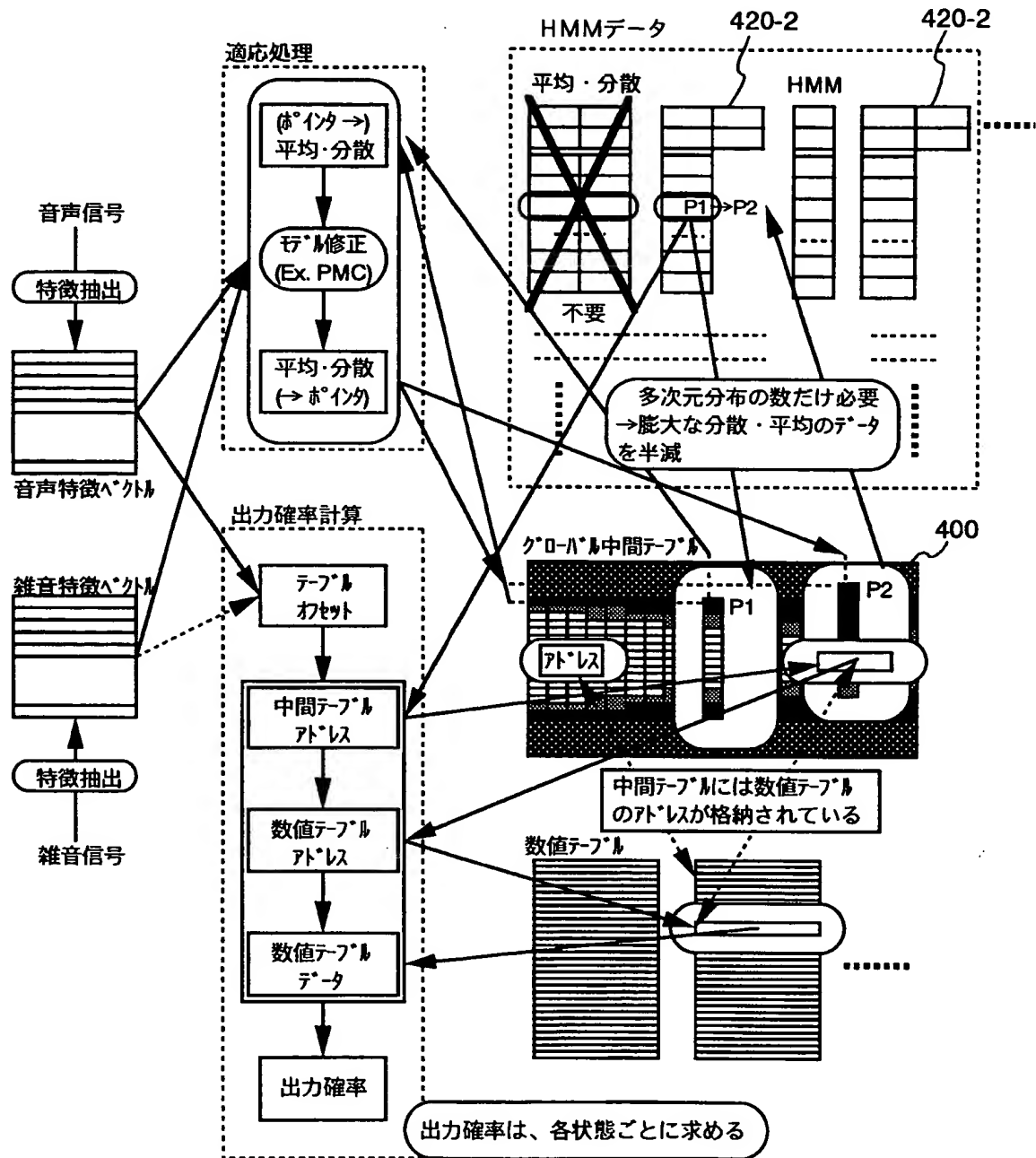
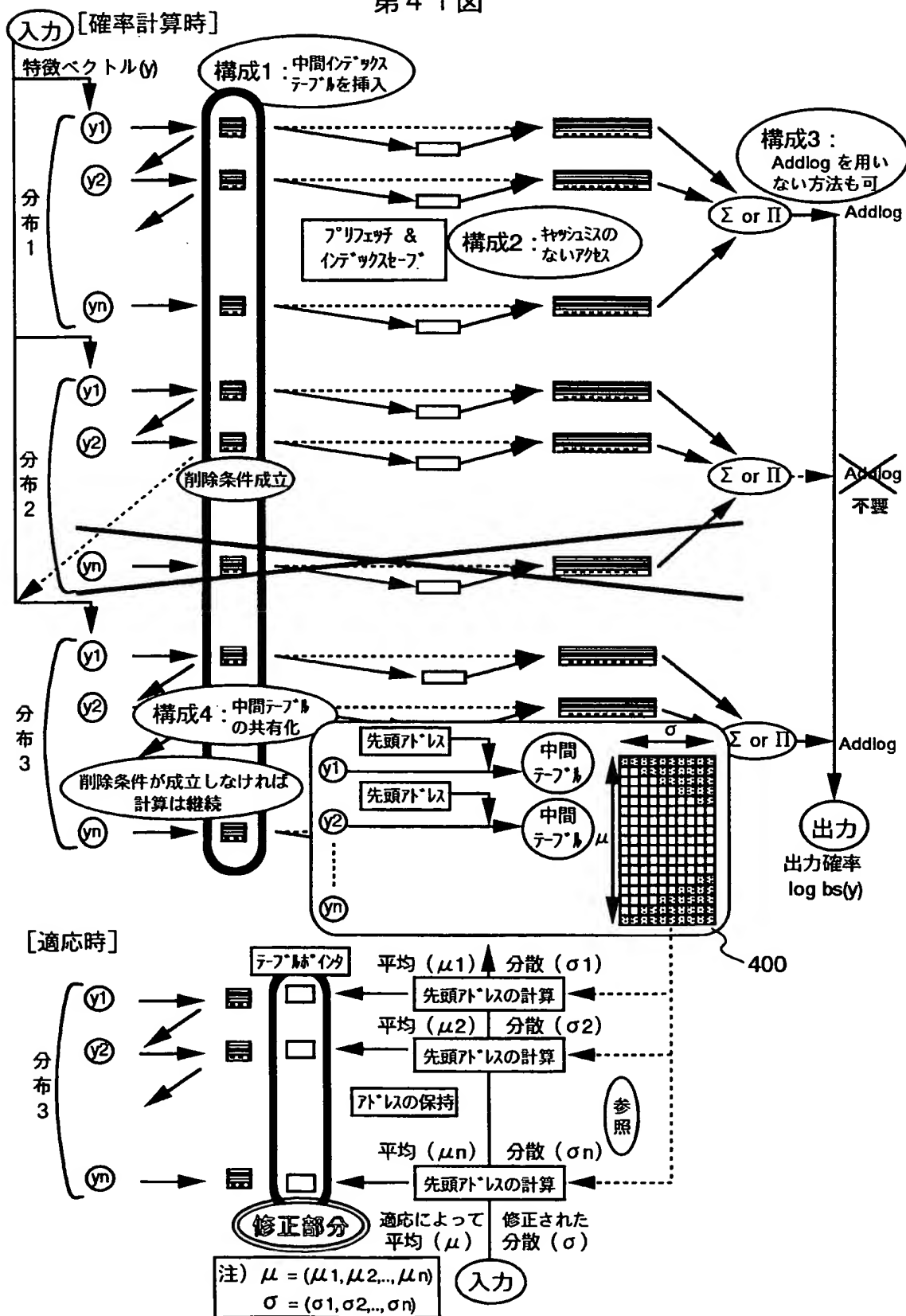


Table Access Means for Probability Calculation using  
Multidimensional Gaussian Distribution

27/28

第41図

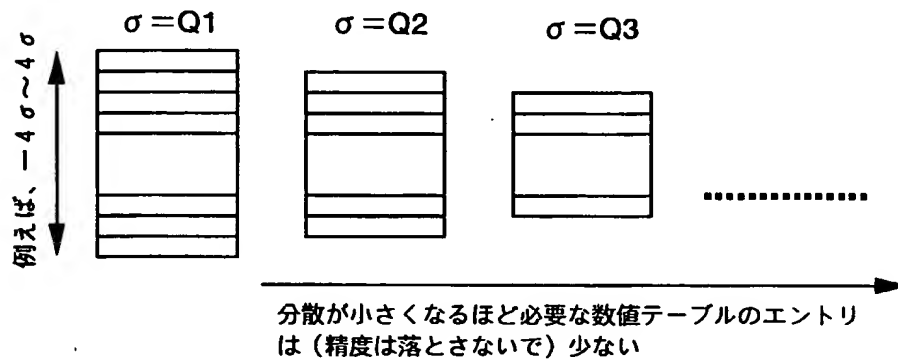


28/28

## 第42図

[浮動小数点演算向き]

$$\log bs(y) = \log \sum_k \omega_k \prod_i \left[ \frac{1}{\sqrt{2\pi \sigma_{ski}}} \exp \left\{ -\frac{(y_i - \mu_{ski})^2}{\sigma_{ski}^2} \right\} \right] \quad R1$$



## 第43図

[整数演算でも対応可能]

$$\log bs(y) = \text{addlog} \left\{ \overset{\text{注)} \quad R3}{\log \omega_k} + \sum_i \left[ \left\{ \frac{-(y_i - \mu_{ski})^2}{\sigma_{ski}^2} \right\} + \log \left\{ \frac{1}{\sqrt{2\pi \sigma_{ski}}} \right\} \right] \right\} \quad R2$$

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP99/00493

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> Int.Cl <sup>6</sup> G10L3/00, 535		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) Int.Cl <sup>6</sup> G10L3/00, 535		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1926-1992 Toroku Jitsuyo Shinan Koho 1993-1999 Kokai Jitsuyo Shinan Koho 1971-1992		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) JICST File (JOIS)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP, 08-248986, A (Satoshi Takahashi, et al.), 27 September, 1996 (27. 09. 96) & EP, 709826, A1 & US, 5794198, A	1-28
A	JP, 08-248985, A (Seiichi Nakagawa), 27 September, 1996 (27. 09. 96) (Family: none)	1-28
A	JP, 09-006382, A (Masaaki Yamada, et al.), 10 January, 1997 (10. 01. 97) & EP, 750273, A3 & US, 5706391, A	1-28
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 30 April, 1999 (30. 04. 99)		Date of mailing of the international search report 18 May, 1999 (18. 05. 99)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer
Facsimile No.		Telephone No.

## 国際調査報告

国際出願番号 PCT/J P 99/00493

## A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl<sup>8</sup> G10L 3/00, 535

## B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl<sup>8</sup> G10L 3/00, 535

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926-1992年

日本国公開実用新案公報 1971-1992年

日本国登録実用新案公報 1993-1999年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICSTファイル (JOIS)

## C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	JP, 08-248986, A (高橋敏 他) 27. 9月. 1996 (27. 09. 96) & EP, 709826, A1 & US, 5794198, A	1-28
A	JP, 08-248985, A (中川聖一) 27. 9月. 1996 (27. 09. 96) (ファミリーなし)	1-28
A	JP, 09-006382, A (山田雅章 他) 10. 1月. 1997 (10. 01. 97) & EP, 750273, A3 & US, 5706391, A	1-28

☐ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

## \* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&amp;」同一パテントファミリー文献

国際調査を完了した日

30. 04. 99

国際調査報告の発送日

18.05.99

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

涌井 智則



5 C

9568

電話番号 03-3501-2695